Using a Novel Game-Like Computerised Measure to Test Executive Functioning in Children

Jennifer Davis

A thesis submitted in partial fulfilment of the requirements of the University of East London for the degree of Professional Doctorate in Clinical Psychology

June 2020

ABSTRACT

Executive function (EF) refers to a group of higher order, complex functions that are crucial for adaptive behaviour. Although it was initially thought that EF did not develop until early adulthood, recent studies have identified that these skills emerge in childhood. Despite this, most tests of EF have been developed for adults and face several challenges including: inconsistencies in the conceptual base, poor specificity, low ecological validity, cultural bias and limited engagement. To address the issue of engagement, several researchers have begun using game-like paradigms.

The present study aimed to further previous research by creating a novel, computerised, EF measure for children that convincingly replicated a game. Three tests were developed with the aim of assessing inhibition, working memory and cognitive flexibility respectively. The novel measure, named Dragon Adventure, was administered to 21 participants aged 11-12 years alongside the following existing measures: D-KEFS Colour-Word Interference test, D-KEFS Trail Making Test, and WNV Spatial Span. Participants rated their enjoyment of each task on a visual-analogue scale. Lastly, teachers completed the Childhood Executive Functioning Inventory (CHEXI) for each child.

Dragon Adventure was found to be ‘enjoyable’, although no more engaging than existing measures of EF. Spearman’s rank correlations revealed moderate-to-large correlations between the novel and established measures, indicating that Dragon Adventure may be successfully measuring EF. Cronbach’s alpha and Spearman-Brown coefficients indicated that the three novel measures had acceptable to good internal consistency. There was a strong association between the Dragon Sequence task and the CHEXI, indicating that this test has good predictive validity.

The results indicate that Dragon Adventure has the potential to be an effective and reliable tool for measuring EF in children. Future research can now be conducted to improve the design of Dragon Adventure, assess engagement, and develop it into a reliable and valid neuropsychological measure.

ACKNOWLEDGEMENTS

I would like to thank all the children who participated in this study for their kind words and creative suggestions on how to develop Dragon Adventure. I would also like to thank the Principal and reception staff at the school for allowing me to run the study and looking after me when I was on site.

Thank you to my supervisor, Dr Matthew Jones-Chesters, for his patience and guidance. You have been a great source of encouragement even at times when I doubted the project would work.

I would like to say a big thank you to my husband, Shivan Davis, who has been an unwavering support throughout the thesis and the three years training. I would also like to thank all my family and friends for all their encouragement and for providing me with much needed respite from my studies.

Finally, a very special thank you to Jaz Thomson who was incredibly generous with his time and patience in helping me learn to use Unity and create Dragon Adventure. I would not have been able to do it without you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

ACC	Anterior cingulate cortex
ADEXI	Adult Executive Functioning Inventory
ADHD	Attention deficit hyperactivity disorder
ASD	Autistic spectrum disorder
BADS-C	Behavioural Assessment of the Dysexecutive Syndrome in Children
BRIEF	Behaviour Rating Inventory of Executive Function
CANTAB	Cambridge Neuropsychological Testing Automated Battery
CFA	Confirmatory factor analyses
CHEXI	Childhood Executive Functioning Inventory
CWIT	Colour-Word Interference test
DANA	Delayed alternation /non-alternation task
DCCS	Dimensional Change Card Sort
D-KEFS	Delis-Kaplan Executive Function System
DLPFC	Dorsolateral prefrontal cortex
EAL	English as an additional language
EF	Executive Function
JBT	Junior Brixton Test
mPFC	Medial prefrontal cortex
NIH Toolbox CB	National Institutes of Health Toolbox Cognition Battery
OFC	Orbitiofrontal cortex
PEL	Primary English language
PFC	Prefrontal cortex
rIFC	Right inferior frontal cortex
SAS	Supervisory Attentional System
SOP	Self-Ordered Pointing
SS	Spatial Span
TEAC-Ch 2	Test of Everyday Attention for Children - Second Edition
TMT	Trail Making Test

TMT – N-L-S	Trail Making Test – Number Letter Sequencing
TMT – N-S	Trail Making Test – Number Sequencing
VLPFC	Ventrolateral prefrontal cortex
WCST	Wisconsin Card Sorting Test
WISC-IV	Wechsler Intelligence Scale for Children - Fourth Edition
WM	Working memory
WNV	Wechsler Nonverbal Scale of Ability

1. INTRODUCTION

The introduction chapter firstly provides an overview of executive functioning and the developmental trajectory in children. It then introduces the literature on the assessment of executive function in children and an example of the neuropsychological tests used. It next discusses the methodological and conceptual challenges in the assessment of executive function in children. Finally, the present study is introduced including the rationale and research questions.

1.1 Executive Function

Executive function (EF) has been described as an umbrella term, referring to a group of cognitive processes that facilitate goal-directed behaviour (Anderson, V., 2002; Goldstein & Naglieri, 2013). However, despite a wealth of research in this area there is no consensus on the definition (Guare, 2014; Jurado & Rosselli, 2007). There is, however, general agreement that the construct of EF refers to higher order, complex functions that are crucial for adaptive behaviour (Best & Miller, 2010; Diamond, 2013; Jurado & Rosselli, 2007; Miyake et al., 2000). One frequently cited definition states that EF enables “independent, purposive, self-directed and self-serving behaviour” (Lezak, Howieson, Bigler & Tranel, 2012, p. 42). EFs have also been described as both ‘cool’, referring to abstract and decontextualized tasks, and ‘hot’, referring to tasks that involve emotion and motivation (Guare, 2014; Hongwanishkul, Happaney, Lee & Zelazo, 2005; Zelazo & Carlson, 2012).

EF is therefore fundamental in successfully completing a wide range of every-day tasks (e.g., planning tasks, setting goals, making decisions, evaluating risks), regulating emotions, and navigating the social world (Diamond, 2013; Fuster, Cole & Tan, 2008; Hofmannm, Schmeichel & Baddeley, 2012; Jurado & Rosselli, 2007). By extension therefore, executive dysfunction refers to problems with planning, self-control, inhibition, and flexibility (Anderson, V., 2001). Individuals who have deficits in EF may have difficulties in planning effectively, initiating and completing tasks, and organisation. Individuals may also experience or express extremes of emotion, for example aggressive outbursts (Hofmannm, Schmeichel & Baddeley, 2012). These difficulties can interrupt social relationships, causing problems in maintaining relationships at home, school or work (Diamond, 2013; Hofmannm, Schmeichel & Baddeley, 2012). Executive dysfunction is hypothesised to be a feature of several developmental disorders including, autistic spectrum disorder (ASD) and attention deficit hyperactivity disorder (ADHD) (Goldstein & Naglieri, 2013).

Historically, research on EF has been associated with examination of patients who have frontal lobe injuries. Researchers noted that patients with these injuries experienced difficulties in attention, self-control, planning, reasoning, and problem-solving (Jurado & Rosselli, 2007; Lezak et al., 2012; Miyake et al., 2000), leading to the hypothesis that EF are linked to the pre-frontal cortex (Jurado & Rosselli, 2007; Miyake et al., 2000). One of the most notable case studies is Phineas Gage, a rail construction foreman who became "disinhibited" and "hyperactive" following a traumatic injury to his left frontal lobe (Ratiu & Talos, 2004). Subsequent neuroimaging studies have also suggested heavy involvement of the frontal lobe, specifically the dorsolateral pre-frontal cortex (DLPFC), ventrolateral prefrontal cortex (VLPFC) and the anterior cingulate cortex (ACC) (Diamond, 2013; Jurado & Rosselli, 2007). EF are not, however, localised to the pre-frontal cortex. They rely on several lower-order functions and appear to be reliant on connections to many areas of the brain, including posterior regions (Jurado & Rosselli, 2007). One widely used model of EF is Miyake et al., (2000) three-factor model: inhibition, updating/working memory and shifting/flexibility (see section 1.2.5 for further discussion of the model). These will now be discussed in turn.

1.1.1 Inhibition

Inhibition refers to the ability to control thoughts, behaviours and emotions in order to prevent a dominant response or one that is internally/externally loaded (Diamond, 2013; Miyake et al., 2000). It can also be considered the adverse or conjoint to 'flexibility' (Diamond, 2013). MacLeod (2007) has argued that the concept of inhibition has been used widely across disciplines (e.g., social psychology, developmental psychology, cognitive psychology), without formal definition, and therefore has become associated with several diverse functions and abilities. There is some agreement that neural inhibition is distinct from cognitive and behavioural inhibition, despite previous research attempting to converge the terms (Bari & Robbins, 2013; MacLeod, 2007). In the field of executive function, voluntary inhibition has been described as a part of cognitive control, comprising both cognitive inhibition (i.e., memories, thoughts, perceptions and emotions) and behaviour inhibition (i.e., response inhibition, deferred gratification and reversal learning) (Bari & Robbins, 2013, p. 51).

Inhibition is strongly associated with activity in the prefrontal cortex (PFC; Bari & Robbins, 2013). Aron, Robbins and Poldrack (2014) argue that inhibition is primarily implemented by the right inferior frontal cortex (rIFC) and is supported by the orbitofrontal cortex (OFC) and the DLPFC which are associated with other executive functions, particularly working memory. Measurement of inhibition has been associated with the classic Stroop Task (Stroop, 1935) and stop-signal tasks, which require participants to inhibit a primed or prepotent response.

In everyday life, good inhibitory control allows individuals to focus on tasks by ignoring distracting stimuli and effectively control their behaviour and emotions. Good inhibitory control is thought to be linked to a good working memory (WM), some hypothesise that this is because they share the same capacity system whereas others suggest that inhibition is a secondary function to WM (Diamond, 2013). Difficulties with inhibition can therefore lead to trouble concentrating on and completing goals, impulsive acts (e.g. being aggressive towards others, shouting out in class or being unable to wait for preferred activities), and difficulties in managing emotions (Hofmannm, Schmeichel & Baddeley, 2012).

1.1.2 Working Memory

The construct of working memory has also attracted several different definitions and the terms “updating” and “working memory” have been used interchangeably (Baddeley, 2003; Miyake et al., 2000). For the purposes of the thesis the term “working memory” (WM) will be used.

WM is the ability to store and code relevant information over short periods of time. Importantly, the function also refers to the ability to manipulate such information, as opposed to passive storage (Miyake et al., 2000). This more complicated skill has been associated with the DLPFC (Baddeley, 2003; Jurado & Rosselli, 2007), whereas simple retention (e.g. digit span forwards) elicits less involvement of the PFC (Diamond, 2013; Miyake et al., 2000).

WM is necessary for any activity that requires retaining relevant information and using such information in relation to a cognitive or behavioural task (Best & Miller, 2010; Miyake et al., 2000). It therefore enables individuals to carry out activities such as: following short instructions in a classroom, weighing up information to make decisions, and remembering and linking concepts to facilitate learning (Diamond, 2013). Individuals who have deficits in WM may struggle to follow instructions (e.g., they may remember the first instruction but not the second) or struggle to make academic progress because information presented earlier has been lost from the WM (Diamond, 2013).

1.1.3 Flexibility

Several terms are used when describing this cognitive process, including “set-shifting”, “switching”, “task switching” and “flexibility” and it is likely to involve several overlapping processes (Dajani & Uddin, 2015). For the purposes of the thesis the term “flexibility” will be used.

Flexibility is the ability to switch between “multiple tasks, operations or mental sets” (Miyake et al., 2000, p. 55), referring to both behavioural change and mental flexibility. This ability requires individuals to disengage from one task and engage in another, despite “proactive interference or negative priming” (Miyake et al., 2000, p. 56). Thus, the ability to switch is linked to inhibition skills. To successfully shift sets, Bissonette, Powell and Roesch (2013, p. 7) propose four tasks that need to be completed:

> “1) forming associations between stimuli, responses and outcomes, 2) detection of errors and conflict between rules, 3) tracking of reward history to determine which responses are no longer valid, 4) enhanced attentional processes to resolve these issues when rules are violated”.

It has been associated with activity in the medial prefrontal cortex (mPFC) and ACC, with the mPFC associated with decision-making and rules and the ACC with conflict-monitoring, although the mechanism behind this activity remains unclear (Bissonette, Powel & Roesch, 2013).

Flexibility is central in the ability to flexibly adapt to a changing environment (Diamond, 2013), for example transitioning successfully between subjects or classes in school. Individuals who have difficulties in flexibility may perseverate, despite information to the contrary (e.g., repeating the same mistake), employ rigid strategies when attempting to problem-solve or struggle to view situations from a different perspective (Dajani & Uddin, 2015; Diamond, 2013).

1.2 Models of Executive Function

Due to the scope of the thesis it is not possible to provide a full review of the models of executive functioning, instead this section will provide a brief overview the main historical and dominant models.

1.2.1 Luria

One of the first people to implicate the frontal lobes in executive functioning was Luria (1973). Luria's model suggests that the brain is divided into three units: lower brain-stem structures, the posterior cerebral cortex and anterior to the central sulcus (MacNeill Horton Jr & Soper, 2008). The third unit most closely resembles executive functioning and is proposed to utilise the frontal lobes in planning and evaluating (Goldstein & Naglieri, 2013).

1.2.2 Baddeley and Hitch

Baddeley and Hitch (1974) were the first to use the term 'executive' in their model of working memory. The model proposes three components: the visuospatial sketchpad which holds and processes visual information; the phonological loop which holds and processes auditory and written information;

and the central executive which is proposed to mediate the other components and govern attention (Baddeley, 2003). Despite it being one of the most widely used models it has been criticised for lacking integration with functional aspects of EF, such as planning and switching between activities (Jurado & Rosselli, 2007).

1.2.3 Supervisory Attentional System

Norman and Shallice (1986) developed the concept of a central executive in their model, the Supervisory Attentional System (SAS) (Baddeley, 2003). They propose that controlled processes (as oppose to automatic processes) require a supervisory system that controls lower order skills to plan, process novel situations, manage danger, correct errors and inhibit inappropriate responses (Chan, Shum, Toulopoulou & Chen, 2008; Jurado & Rosselli, 2007).

1.2.4 Tripartite model

Stuss and Benson's (1986) model moved away from a unitary model of EF and proposed that attention and EF are managed by three interrelated components. The anterior reticular activating system and the diffuse projection system are hypothesised to monitor alertness, while the fronto-thalamic gating system manages higher level processes such as planning, inhibition and self-monitoring. There is also an increased focus on the neural underpinnings of the aspects of EF (Chan et al., 2008).

1.2.5 Miyake model

One of the most influential models is Miyake et al., (2000) model of EF. Miyake et al., (2000) administered the Wisconsin Card Sorting Test (WCST), Tower of Hanoi, random number generation, operation span, and dual tasking test to 137 undergraduates. Using confirmatory factor analysis, they propose three inter-related factors: inhibition, updating/working memory and shifting/flexibility.

Several subsequent research papers have also suggested the same broad three factors across the age-span (e.g., Fisk & Sharp, 2004; Karr et al., 2018;

Lehto et al., 2013) and Karr et al. (2018) systematic review concluded that the most commonly accepted model for school-aged children was a three-factor model. Since the original paper, Miyake and Friedman (2012) and Friedman and Miyake (2017) have published further evidence and versions of the model and Diamond (2013) concludes that there is a general agreement in support of these three factors. As such, the current thesis has adopted the Miyake model as a prominent framework to understand EF.

It should be noted that the model does not suggest that these are the only domains of EF. Diamond (2013) suggests that these three core EF domains are utilised to establish higher-order EF skills such as planning, problem solving and self-regulation. Although not focused on in the present thesis, these additional EF skills are relevant when considering the relative strengths and weaknesses of children supported in clinical psychology services.

1.3 Development of EF

The development of EF is a large and growing area of study, so only the main germane themes in the literature can be addressed here. The assessment of EF in children has been neglected due an assumption that EF does not mature until adulthood (Anderson, P., 2002). Recent research has challenged this, revealing that several domains of EF begin to develop in childhood (Anderson, V., 2002). The pattern of development of these skills is uneven, with periods of rapid improvement and periods of gradual change (Anderson, V., 2002; Best & Miller, 2010).

It is hypothesised that the development of EF is aligned with the brain's development, particularly of the PFC (Anderson, V., 2002). The brain undergoes its most rapid and dramatic development in utero, however maturation continues into late adolescence and adulthood (Casey, Giedd & Thomas, 2000). The development of the PFC is more protracted than other brain regions, maturing in later adolescence (Best & Miller, 2010). Development consists of progressive (e.g., myelination and synaptogenesis) and regressive changes (e.g., synaptic pruning) resulting in extensive neural networks across the brain (Best & Miller, 2010).

The domains of EF do not appear to develop at the same rate, instead emerging at different ages (Best & Miller, 2010). Overall, they appear to be relatively *established* by 11-13 years, *improve* across adolescence and *mature* in early adulthood (Anderson, P., 2002; Goldstein & Naglieri, 2013). There is no unified model of the development of EF in typically developing young people and studies investigating the development of EF in children are constrained by the validity and suitability of EF tests for children (see section 1.5.2 Challenges in assessment). Nevertheless, several researchers have begun to draw together existing research (Anderson, V., 2002; Best & Miller, 2010; Diamond, 2006; Garon, Bryson & Smith, 2008). See tables 1, 2, and 3 for a summary of the development of each domain across the developmental stages (Piaget, 1936).

Inhibition and WM skills appear to emerge before flexibility, perhaps because the ability to set-shift is, in part, dependent on inhibition and WM (Best & Miller, 2010). The precursors to WM, the Short-Term Store, develop in infancy, with 6-month-olds able to hold information for very short periods of time (Best & Miller, 2010; Garon et al., 2008). There appears to be a mostly linear improvement in WM ability from pre-school years until adolescence (Best & Miller, 2010; Garon et al., 2008), with performance on simple WM tasks reaching adult levels by approximately 8 years and performance on complex WM tasks continuing to gradually improve into adulthood (Best & Miller, 2010).

Inhibition skills also begin to emerge by 6 months, with infants able to succeed on detour reaching and “don’t” paradigm tasks (Diamond, 2006; Garon et al., 2008). Inhibition appears to develop more rapidly during certain periods. There is a significant increase in preschool years, where there is a marked improvement on inhibition tasks, particularly those that involve WM (Garon et al., 2008). Following this, development appears more subtle, with improvements in tasks that require stopping an already initiated response and increasingly complex tasks that demand both inhibition and WM (Best & Miller, 2010; Garon et al., 2008).

Fundamental flexibility skills are apparent between the ages of 3-4 years, where children can switch between two simple tasks (Anderson, V., 2002). Between the ages of 4 to 7, children begin to succeed on tasks that involve unexpected shifts (Best & Miller, 2010), for example the WCST (Diamond, 2006). Improvement in flexibility continues to develop throughout later childhood and adolescence until reaching maturity at 20+ years (Anderson, V., 2002; Diamond, 2006). The speed-accuracy trade-off is one of the later components to develop, with adults preferring accuracy over speed compared to adolescents (Best & Miller, 2010). This is likely linked to the different developmental trajectories of 'hot' and 'cool' EFs, with 'cool' EFs becoming established earlier in development than 'hot' EFs which are seen to develop throughout adolescence (Zelazo & Carlson, 2012).

As stated above, the findings about the development of EF are inevitably constrained by the complexity of assessing EF in children. Developing appropriate measures is complicated by the fact that it has only recently been considered to be an issue, and because it must be developmentally appropriate considering the child's overall cognitive development. The next section details a literature search that focuses on both the assessment tools and challenges in measuring EF in children.

Table 1. Development of Inhibition

	0-2 Sensorimotor stage	2-7 Pre-operational stage	7-11 Concrete operational stage	11+ formal operational stage
Inhibition	**6-12 months:** Children begin to succeed on the detour reaching task which requires the ability to hold a goal in mind, inhibiting the reaction to go straight to an object and plan a route (Diamond, 2006; Garon et al., 2008). Infants can inhibit a rewarding behaviour in the "don't" paradigm 40% of the time at 8 months (Garon et al., 2008).	**2-3 years:** Children can inhibit a rewarding behaviour in the "don't" paradigm 78% of the time at 22 months and 90% of the time at 33 months (Garon et al., 2008). The length of time children are able to wait in the delayed gratification task increases from 2 years to 4 years (Garon et al., 2008). **3-5 years:** Marked improvement in complex inhibition tasks that require WM (Garon et al., 2008) and a reduction in perseveration (e.g., success with day-night Stroop task) (Gerstadt, Hong, Diamond, 1994) and Luria's hand game (Hughes, 1998) (Best & Miller, 2010; Diamond, 2006). At age 4.5 years children can successfully inhibit incorrect responses on go-no-go task (Diamond, 2006; Livesey & Morgan, 1991). Increasing the complexity of the rules in conflict tasks decreases performance, for example 5-6 years olds struggle with the advanced DCCS (Best & Miller, 2010; Carlson, 2005).	**8-12 years:** The level of improvement in inhibition is contended. Some studies find little improvement beyond 8 years (Best & Miller, 2010; Klenberg, Korkman, Lahti-Nuuttila, 2001; Lehto, Juujärvi, Kooistra, Pulkkinen, 2003). However, there is evidence that performance on go/no-go tasks and continuous performance tasks improve from age 9 to adolescence, particularly in tasks that require stopping an already initiated response (Best & Miller, 2010).	Subtle improvements in performance on complex inhibition and WM tasks (Best & Miller, 2010).

Table 2. Development of Working Memory

	0-2 Sensorimotor stage	2-7 Pre-operational stage	7-11 Concrete operational stage	11+ formal operational stage
Working Memory	**6-12 months:** The ability and length of time an infant can hold a representation in mind develops, from a few seconds at 6 months to 10+ seconds at 12 months (Garon et al., 2008).	**2 years:** Children begin to succeed at the invisible displacement task (Garon et al., 2008). **3-7 years:** Children aged 3-5 years become more accurate on the self-ordered pointing task (Garon et al., 2008). Number of items recalled in digit-word and object-spatial span improves, at approximately 4 items for 5-year-olds (Garo et al., 2008) There is a marked improvement in complex span tasks (Diamond, 2006), with average number of items improving from 1.58 at age 3 to 2.88 at age 5 (Garon et al., 2008). Performance on simple WM tasks improves until 7-8 years where performance is equivalent to adolescents and young adults (Best & Miller, 2010).	**8-11 years:** Number of items recalled in digit-word and object-spatial span improves to approximately 14 items at age 11 (Garon et al., 2008). Gradual improvement in complex span tasks (Diamond, 2006). Gathercole, Pickering, Ambridge and Wearing (2004) found linear increase in performance across WM tasks from the age of 4 to 15.	The executive WM continues to develop and performance on complex WM tasks improves throughout adolescence and gradual improvement seen into adulthood (Best & Miller, 2010).

Table 3. Development of Flexibility

	0-2 Sensorimotor stage	2-7 Pre-operational stage	7-11 Concrete operational stage	11+ formal operational stage
Flexibility		**3-4 years:** Children begin to be able to switch between two simple tasks (Anderson, V., 2002) and where rules are embedded in a story (Hughes, 1998). The ability to shift appears contingent on the development of inhibition and WM, and is therefore less developed (Best & Miller, 2010). **4-5 years:** Marked improvement in cognitive flexibility e.g. success on DCCS task (Diamond, 2006; Zelazo, Reznick & Piñon, 1995). Although reducing inhibitory demand in the task increases success in 3-year-olds (Diamond, 2006). **6-7 year:** Aged 5-6 children show success at stage 7 of the intradimensional/ extradimensional self-shifting task, using feedback from previous trials (Luciana & Nelson, 1998). Emerging success on the Wisconsin Card Sorting Test (WCST) (Diamond, 2006; Heaton, 1981).	**7-9 years:** Children can manage multi-dimensional switching tasks (Anderson, V., 2002). **7-11 years:** Ability to switch back and forth in tasks improves, although speed and accuracy remain poorer than adult levels (Diamond, 2006). Improved performance on WCST (Diamond, 2006; Heaton, 1981). Performance on flexibility tasks continues to improve and children become better able to manage unexpected shifts (Best & Miller, 2010).	Cognitive flexibility and switching continues to improve and mature until 20+ years (Anderson, V., 2002; Diamond, 2006). Peak performance on WCST (Diamond, 2006; Heaton, 1981). Performance on some computerised flexibility tasks level out at 15 years old (Huizinga, Dolan & Molen, 2006). Speed-accuracy trade-off develops, with adults preferring accuracy over speed compared to adolescents (Best & Miller, 2010). Increase in success rates through to stage 9 of the intradimensional/ extradimensional self-shifting task in young adults (Luciana & Nelson, 1998).

1.4 Literature Review Process

1.4.1 Method

A literature search of papers relating to the measurement of EF in children was performed using Academic Search Complete, CINAL Plus, Child Development and Adolescent Studies, PsychInfo and Scopus databases. The search terms used were "executive function" (and exploded terms) with “assess*”, “cognitive”, “cognition”, “neuropsych*” and “game” in different combinations. Where possible, the results were restricted by age.

1.4.2 Inclusion and exclusion criteria

Due to the scope of the literature available, the search was limited to papers that investigated the measurement of executive function in typically developing school-aged children. Only papers that focused on performance-based and behavioural measures were included. Both papers describing measures and those discussing the complexities of measuring EF in children were included. Lastly, only papers published in English were included.

1.4.3 Search results

Figure 1 shows the PRISMA Flow Diagram (Moher, Liberati, Tetzlaff, & Altman, 2009) describing the papers identified, screened and selected for the review. The initial search identified 20,215 papers. After removing duplicates, the number of papers remaining was 12,422. Due to the large amount of papers identified, papers with key words in the title that indicated they did not meet the inclusion criteria were removed. The key words used for removing papers were:

> “addiction”, “ADHD”, “alcohol”, “autism”, “bilingual”, “bipolar”, “brain injury”, “cancer”, “cannabis”, “cardiac”, “cocaine”, “deletion syndrome”, “dementia”, “depression”, “diabetes”, “disorder”, “down syndrome”, “dyslexia”, “epilepsy”, “forensic”, “homeless”, “intellectual disabilit*”, “language impairment”, “mild cognitive impairment”, “multiple sclerosis”, “obesity”, “older adults”, “opiate”,

“preschool”, “preterm”, “psychosis”, “PTSD”, “schizophrenia”, “schizophrenic”, “schizotypal”, “stroke” and “trauma”.

Following this, the number of papers remaining was 3,372. The titles and, where necessary, the abstracts were reviewed to identify studies that met the inclusion criteria. The full text of the remaining 77 papers were then read and 47 selected to be included in the review. Due to the large number of papers identified in the initial search strategy, using fewer databases and narrower search terms may have been useful to limit the search.

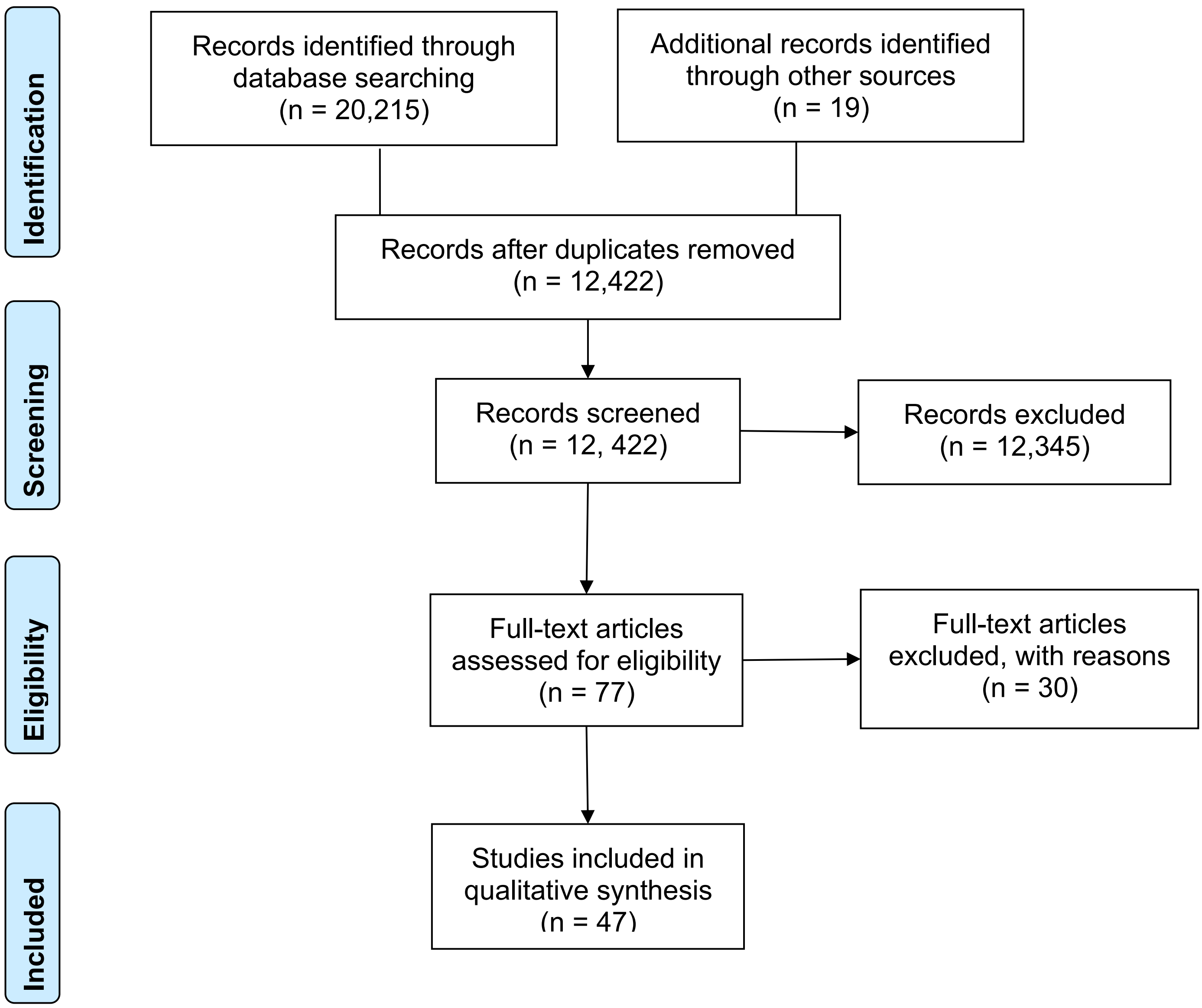


Figure 1. PRISMA (Moher et al., 2009) Flow Diagram of Article Selection Process.

1.5 Literature Review

This section details the qualitative synthesis from the 47 papers identified in the literature search. It will first discuss the methods of assessing EF in children identified in the review. It will then describe the challenges in accurately assessing EF in children: definition of EF, developmental needs, specificity, ecological validity, cultural bias and engagement. It will also describe the approaches that have utilised a game-like protocol in the measurement of EF in children. Lastly, the justification and research questions for the current study will be presented.

1.5.1 Assessing EF

To test EF, measures need to be novel and complex (Anderson, V., 2001) so that they require the child to develop new strategies or schemas, rather than complete them automatically, on the basis of existing knowledge. As discussed above, EF is central to everyday living and deficits in this area can have a significant impact on functioning. Clinicians need to be able to accurately measure a child's EF in order to understand their areas of relative strength and difficulty as well as informing intervention (Anderson, V., 2001). For example, in child and adolescent mental health services (CAMHS) children can present with a range of difficulties, accurate assessment tools are vital to understand whether cognitive factors should form part of the formulation. It is not possible to describe all the measures used in the assessment of EF in children due to the large number of instruments used. Below is a summary of the mains paradigms, individual measures and assessment batteries identified in the literature search. Although they have been broadly categorised into inhibition, working memory and flexibility it should be noted that all measures require overlapping skills from different domains.

1.5.1.1 Inhibition

Go-No-Go paradigms have been used to measure inhibition in children. They typically measure the ability to inhibit a dominant response by requiring the participant to not respond to a stimuli that has a prepotent response (Archibald & Kerns, 1999). For example, the participant needs to press a button every time a green stimulus is shown (frequent) and then do nothing when a red stimulus is shown (infrequent). Performance is measured by error rate, where fewer errors indicate better inhibitory skills. Archibald and Kerns (1999) provide normative data for children aged 7-12 years.

The classic Stroop task requires individuals to read aloud the colour ink a word is written in whilst ignoring what the word says (e.g., the colour ink is red and the word written is blue) and measures the ability to ignore salient but irrelevant stimuli (Archibald & Kerns, 1999). As it is a word-based task, performance depends on literacy levels. Several variants of the Stroop task have been used with children. For example, the day-night task (Gerstadt, Hong & Diamond, 1994) is frequently used with pre-schoolers and requires the child to say ‘day’ when they see a picture of a moon and ‘night’ when they see a picture of a sun, but has been found too simple for older children. Lagattuta, Sayfan, and Monsour (2011) developed a happy-sad task in which children are required to say ‘happy’ when they see a sad face and ‘sad’ when they see a happy face. They tested the new measure with 350 participants aged 4-27 and found it to be sensitive enough to measure performance in both children and adults.

The NEPSY II (Korkman, Kirk & Kemp, 2007) is battery designed to test cognition in children aged 3-16 years and contains the ‘Inhibition’ subtest. Children are shown shapes and arrows, each either black or white. The child is instructed to name either the shape or direction, inhibiting the prepotent response. Standardised data for the entire NESPY II has been collected from 1200 UK children and Korkman et al. (2007) found it to have low to moderate convergent validity with the Delis-Kaplan Executive Function System (D-KEFS; Delis, Kaplan & Kramer, 2001).

The Cambridge Neuropsychological Testing Automated Battery (CANTAB; Cambridge Cognition, 2006; Luciana, 2003) is a computerised battery designed to test cognitive function, including executive functioning, across the lifespan. Luciana and Nelson (2002) measured performance in 4-12-year-old children to start developing normative data. It contains the Multitasking Test which requires participants to manage conflicting information. Participants need to select a right or left button according to an arrow displayed on the screen. These arrows can be congruent or incongruent, hence demanding the participant ignore task-irrelevant information.

The National Institutes of Health Toolbox Cognition Battery (NIH Toolbox CB; Bauer & Zelazo, 2013; Weintraub et al., 2013) is designed to be a brief measure of cognition that can be used thoughtout the life span. It includes the Flanker Inhibitory Control and Attention Test. It is adapted from the Eriksen flanker task (Eriksen & Eriksen, 1974). Participants need to indicate the orientation of a stimulus (left or right) whilst inhibiting attention to incongruent stimuli (flankers, e.g., arrows). Zelazo et al. (2013) found that the EF measures within the toolbox (Dimensional Change Card Sort and Flanker Inhibitory Control test) were sensitive and reliable for children aged 3-15 years.

The Delis-Kaplan Executive Function System (D-KEFS; Delis et al., 2001) is a standardised assessment battery of EF for people aged 8-89 years. Regarding inhibition, it contains the Colour-Word Interference Test (an adaption of the classic Stroop task) and the Tower Test (similar to the Tower of Hanoi). The D-KEFS has been standardised on a sample of 1750 individuals (Delis et al., 2001), Latzman and Markon (2010) used this data and a further sample of males aged 11-16 to conclude a three-factor solution labelled Conceptual Flexibility, Monitoring and Inhibition. Despite it being widely used to measure EF in children, Fisher (2006) found that it may lack validity for this age group, although they only administered the battery to 28 children 8-12 years.

1.5.1.2 Working memory

A wide variety of both verbal and visual span tasks exist for the measurement of EF in both children and adults (Henry & Bettenay, 2010). Digit span tasks that include a backwards component are frequently included in larger test batteries (Henry & Bettenay, 2010). These require the child to remember a sequence of numbers and then repeat them backwards. Spatial spans tasks (e.g., Wechsler & Naglieri, 2006) are a similar paradigm for the visual domain, in which the administrator points to different cubes and the child needs to remember the sequence and then repeat backwards.

The CANTAB (Cambridge Cognition, 2006) contains a digital Spatial Span test, very similar to other spatial span measures described above. It also includes a Spatial Working Memory test, during which participants are shown coloured squares and must select boxes until they ‘find’ a yellow token. Difficulty is increased by increasing the number of squares displayed. The CANTAB One Touch Stocking of Cambridge test contains a working memory subdomain. Based on the Tower of Hanoi test, participants are required to move coloured balls from one display to another. In the working memory subdomain, the participant is shown a problem, then must calculate the number of moves required to move the coloured balls and select the appropriate box to indicate how many moves are required.

The NIH Toolbox BC (Bauer & Zelazo, 2013; Weintraub et al., 2013) includes the List Sorting Working Memory Test. Participants are shown familiar stimuli in varying sizes (e.g., pictures of animals or food). They are required to remember the stimuli, sequence them in order of size and then verbally report the items in the correct order. The NEPSY II (Korkman et al., 2007) contains the Word List Interference test, designed to measure verbal working memory. The child is presented with two lists of words and asked to repeat the list after each has been presented. They are then asked to recall each list in the order they were presented.

The Self-Ordered Pointing (SOP) test was developed by Petrides and Milner (1982) and has subsequently been investigated as a measure for children. It requires children to remember familiar and abstract designs. The child needs to point to each design, which are presented on separate cards in different locations, until all designs have been pointed to. Archibald and Kerns (1999) provided normative data on 89 children aged 7-12 and Cragg and Nation (2007) found the SOP was a sensitive measure of working memory in children, with performance improving with age.

Archibald and Kerns (1999) created an adapted version of the Delayed alternation /non-alternation task (DANA) for children and provided normative data. The computerised task requires children ‘throw’ basketballs into hoops. To succeed on the delayed alternation block, children need to choose the opposite hoop to the one previously chosen. The delayed non-alteration block is then presented without warning, in which the child must always chose the same hoop. The delay period, since the last trial, can be increased to increase the demand on WM.

1.5.1.3 Flexibility

Set-shifting paradigms are commonly used to measure cognitive flexibility. Tasks typically involve the participant learning a pattern or response, which is then switched to an alternative target pattern or response. Performance is usually measured through error rate and speed.

The Trail Making Test (TMT) requires individuals to sequence a series of letter and numbers, switching between the two types of stimuli. The child version of the TMT (Reitan, 1992) requires children to first sequence the numbers 1-15 and then sequence the numbers 1-18 and letters A-H. Using the comprehensive trail making test (Reynolds, 2002), Riccio, Kahn, Yoon, Reynolds and Bonura (2011) conducted a confirmatory factor analysis on the performance of 557 children aged 8-18 years and concluded a two factor model (hypothesised to be sequencing and shifting) best fit the data. The D-KEFS (Delis et al., 2001) also contains a version of the Trail Making Test as well as the Verbal Fluency Test which requires participants to generate words associated with a letter, category and finally alternating between two categories.

The Shape Trails Test (Zhao et al., 2013) is designed to be more culturally fair, due to the removal of the alphabet as a stimuli. Chan and Morgan (2018) adapted the adult Shape Trails Test for children. Part A requires children to sequence the numbers 1-15. Part B requires children to sequence the numbers 1-7 and 1-8 alternating between numbers enclosed in circles and numbers enclosed in squares. They administered the test to 68 six to nine-year-old children and concluded that it is a valid measure of cognitive flexibility in this age group.

In the Wisconsin Card Sorting Test (WCST; Berg, 1948) individuals are required to categorise different cards according to the colour and shape of the stimuli on each card. The rule periodically changes, requiring the individual to shift set and categorise according to the new rule. Chelune and Baer (1986) developed normative data for the WCST in a group of 105 school-aged children and found that performance reached adult levels by age 10. They note the conceptual issue of using adult measures in children but conclude that the WCST can be useful in the assessment of children if embedded in an understanding of child neurodevelopment. Bujoreanu and Willis (2008) administered the standard and an adapted version (number sorting criteria altered) of the WCST to 196 children aged 6-19 years. They found that performance and number of categories completed increased with age.

Another variation of the card sorting task is included in the NIH Toolbox BC (Bauer & Zelazo, 2013; Weintraub et al., 2013) is the Dimensional Change Card Sort Test. This requires children to sort cards based on colour and shape. Switching the sorting rules evokes flexibility. As stated above, Zelazo et al. (2013) found this test to be sensitive and reliable for children.

The Test of Everyday Attention for Children, second edition (TEA-Ch 2; Manley et al., 2016) also contains a variation of the card sorting task. It is suitable for children aged 8-15 years and utilises a comic-based format to increase engagement. The Reds and Blues, Bags and Shoes test requires children to sort the four stimuli according to colour and location on a character (i.e., hand or foot). Normative data has been collected from 621 children (Manley et al., 2016).

The Behavioural Assessment of the Dysexecutive Syndrome in Children (BADS-C; Emslie, Wilson, Burden, Nimmo-Smith & Wilson, 2003) contains a variation of the card sort task, named the Playing Cards Test, for children aged 7 to 16 years. It also contains the Water test, Key Search Test, Zoo Map Test, and Six Part Test, to measure planning, flexibility, verbal and visuo-spatial skills. These novel tests can be difficult to compare with other measures due to limited research on what domain of EF each test measures and have lacked test re-test reliability (Henry & Bettenay, 2010).

In the Contingency Naming Test (Taylor, Albo, Phebus, Sachs & Bierl, 1987) children are instructed to identify either the colour or shape of stimuli according to different rules. Trial one requires the child to identify the colour and trial two the shape. In the third trial, the child must alternate between colour and shape depending on whether internal and external shapes match. The final trial requires the same as trial three, but an arrow indicates when the child should reverse the rule. Anderson, Anderson, Northam and Taylor (2000) provided normative data from 381 children aged 7-15 years and concluded it is a sensitive measure for reactive flexibility in children.

Harvey, Rose, Jonsson and Lask (2016) administered the adult Brixton Spatial Anticipation Test to 72 female participants aged 11-17 years to assess the validity of using the test for children and adolescents. They found no significant difference on performance between age groups, contrary to literature that suggests executive function improves with age. The authors acknowledged that the female only and small sample size was a significant limitation.

In the Junior Brixton Test (JBT; Senturk, Yeniceri, Alp, & Altan-Atalay, 2014) children are presented with cards with an array of turtles (one green and nine grey) in differing positions. The aim is to identify the position of the green turtle on the upcoming card. Senturk et al., (2014) administered the JBT and the WCST in 121 6-8-year-olds. They argue that the JBT paradigm was more child-friendly and more sensitive at discriminating between variations in perseverative errors than the WCST and conclude it is a useful measure when used with other cognitive tests.

Klimkeit, Mattingley, Sheppard, Farrow and Bradshaw (2004) used a novel selective reaching task to assess executive function in 7-12-year-old children. In the ‘maintain-set’ condition the task required children to respond to green target lights and ignore red distractor lights by pressing keys. In the ‘change-set’ condition the target light switched between red and green, indicated by the colour of an initial fixation light. 8-year-olds made significantly more errors than 10 and 12-year-olds and younger children took longer to respond than older children. They concluded that the task was sensitive to measuring flexibility, attention and response inhibition.

The CANTAB (Cambridge Cognition, 2006) contains the Intra-Extra Dimensional Set Shift task to measure flexibility. Participants are presented with pink shapes and white lines. Participants must learn the rule that determines which stimulus is correct (e.g. pink shapes are relevant vs white lines are relevant). These rules shift, requiring participants to update their strategy.

The NEPSY II (Korkman et al., 2007) contains the Auditory Attention and Response Set. The response set is designed to measure the child’s ability to shift set. The child listens to a series of words and touches the appropriate response stimuli upon hearing a target word. The response set assesses the child’s ability to shift to a new set and inhibit previously learned response.

1.5.2 Challenges in assessment

Numerous authors discussed the significant challenges in measuring EF in children. These are discussed below.

1.5.2.1 Definition and construct of EF

As described above, there is little consensus on a definition of EF, models of EF or on the number of components. Similarly, the tests used to measure EF use a large variety of terms to describe the skill they claim to measure. Developing tools that can accurately measure EF is extremely challenging due to the inconsistency in the conceptual base (Fahy, 2014).

Willoughby, Holochwost, Blanton and Blair (2014) call attention to the fact that performance-based tests used to assess EF are typically only weakly correlated, with even tests that aim to measure the same dimension of EF often weakly correlated. Pureza, Jacobsen, Oliveira and Fonseca (2011) compared several tests of EF and found a large variation in the relationship between the tests, ranging from .3 to .7. They argue that this is a significant limitation for those attempting to use confirmatory factor analyses (CFA) to understand the construct of EF. Wiebe and McFall (2014) note that using a large range of test can help to overcome this limitation, but also that this becomes increasingly impractical when working with children. Willoughby et al. (2014) also question whether performance-based tasks are causal (formative) or an effect (reflective) of the construct of EF and Willoughby and Blair (2016) conclude that many EF tests are better conceptualised as formative. They urge researchers to consider the implications using CFA to model the concept of EF although Wiebe and McFall (2014) argue that a broad measurement approach is needed in both scenarios.

Lehto et al., (2003) found a similar three-factor structure of EF in children aged 8-13 years, to Miyake et al. (2000) model in young adults. Wasserman and Wasserman (2013) conclude that these three factors appear relevant for children but highlight their interconnectedness and question the utility of trying to sperate them entirely.

1.5.2.2 Specificity

In addition to the problem of definition, tasks that measure EF are required to be complex, because of this they necessitate the use of lower-order functions such as attention and comprehension (Anderson, V., 2001; Salimpoor & Desrocher, 2006; Wasserman & Wasswerman, 2013). It has therefore been challenging to identify the role of specific components of EF (e.g., flexibility, inhibition or working memory) and whether any difficulties on a task are due to problems EF directly or with one or more lower-order functions. To illustrate this challenge, Wasserman and Wasserman (2013) point to the WCST, which at various points has been claimed to measure "set-shifting…, auditory working memory, verbal fluence, and attention" (p. 89).

To address the challenge of specificity, Reynolds and MacNeill Horton Jr (2008) recommend using a variety of tasks that require differing levels of motor and verbal skills, to minimise the effects of lower-order skills obscuring the measurement of EF.

1.5.2.3 Developmental considerations

Most tests of EF were originally designed for use in adult populations and there is a significant lack of normative data for younger people and children (Anderson, P., 2002; Anderson, V., 2001; Anderson, V., 2002; Archibald & Kerns, 1999). If thorough normative data is not available for children, clinicians will have significant difficulty in interpreting potential deficits that the child may have, for example those who have experienced a brain injury (Anderson, V., 2001). Tests created for adults are also based on several assumptions that may not hold true for children. One issue is the assumption that localised skills and/or deficits that are seen in adults are present in children as oppose to a global EF skill (Anderson, V., 2001). Even tests that have been specifically designed for children will struggle to be developmentally appropriate for both younger and older children due to rapid developmental changes (Anderson, V., 2002).

1.5.2.4 Ecological validity

The existing measures have also been criticised for lacking ecological validity, as differences between performance on traditional measures and ability in everyday tasks are often noted (Anderson, V., 2001; Silver, 2014). Traditional tests typically occur in a quiet room with a single examiner, this may fail to elicit difficulties that become apparent in contexts that place greater demand on EF and mean that difficulties with EF go undetected (Anderson, P., 2002, Anderson, V., 2002; Fernández, González-Castro, Areces, Cueli, & Pérez, 2014; Salimpoor & Desrocher, 2006; Silver, 2014; Wallisch, Little, Dean & Dunn, 2018; Wasserman & Wasserman, 2013). Wallisch et al. (2018) conducted a scoping review of ecological validity in EF measures for children. They also found that many measures of EF in children occurred within structured environments and even those that aimed to replicate real-life scenarios did not accurately reflect real-life tasks.

To address the issue of ecological validity, many clinicians use questionnaires such as the Behaviour Rating Inventory of Executive Function (BRIEF, Gioia, Isquith, Guy & Kenworthy, 2000), however these have been found to be poorly correlated with performance on objective neuropsychological tests (Fernández et al., 2014; Silver, 2014; Toplak, West & Stanovich, 2013). Where there is a discrepancy, clinicians need to try and understand whether these differences are because performance-based tests are not sensitive enough to detect EF difficulties, or whether the questionnaires are measuring an entirely different construct. Finally, Silver (2014) noted that computer-based assessments provide an opportunity to improve ecological validity by providing life-like paradigms of challenges that are likely to be faced in the ‘real world’.

1.5.2.5 Hot and cool tests

There is increasing interest in the distinction between ‘hot’ and ‘cool’ EF and when these capacities develop. There is concern, however, that the current measures may be inadequate at distinguishing between the two (Welsh & Peterson, 2014). Welsh and Peterson (2014) argue that tests that are traditionally conceptualised at ‘cool’, in fact involve a significant emotional component for the child, implying that ‘hot’ EF are also being utilised. They also note that the level of arousal experienced during a task is likely to vary between children, so that a task may be ‘hotter’ for one child and ‘cooler’ for another. They propose that the ‘heat’ of single tasks could be manipulated (e.g., by increasing or decreasing rewards) to assess the relative development of ‘hot’ and ‘cool’ EF. McCoy (2019) note that ‘hot’ tasks may be more ecologically valid, however they are still unlikely to reflect the complexity of real-world experiences.

1.5.2.6 Context

Wallisch et al. (2018) found that few measures of EF attempted to understand the contextual factors related to EF performance. There is also widespread criticism of standardised tests as being culturally biased (Fernández & Abe, 2018). There was very little discussion of cultural bias being a challenge in the assessment of EF in children, which is perhaps reflective of a wider neglect of this subject. Nevertheless, Fernández and Abe (2018) note several areas of bias that apply to neuropsychological testing more generally: the assumption that a construct that may be valid in one cultural group is valid in another; method and administration bias, whereby study samples favour a particular group or study instruments that are more familiar to a particular group; differential item functioning (DIF), where the same item, within a large construct, has different meanings in different cultures; and language bias, both English as an additional language and the assumption that tests can be translated and retain the same meaning.

1.5.2.7 Engagement

Researchers also highlight that children may lack interest in existing measures of EF (Anderson, V., 2001). Motivation is key in initiating behaviour and is therefore important in EF (Guare, 2014). As noted above, most tasks of EF are likely to involve the use of both ‘hot’ and ‘cold’ EFs (Hongwanishkul et al., 2005), if children are not engaged and motivated to persevere and attend to the test, their performance is unlikely to be reflective of their true ability. Tests of EF, therefore, need to be engaging for children.

1.5.3 Game-like procedure

A game can be defined as “an activity or sport usually involving skill, knowledge, or chance, in which you follow fixed rules and try to win against an opponent or to solve a puzzle” (Collins, 2020). It can also be considered a type of ‘play’, which is thought to be a part of development for almost all children (Cohen, 2006). Regarding video-game theory, Crawford (2012, pg. 67-96) details several key components of video games:

- **Rules** of play and underlying logic.
- **Interactivity** between the player and the game experience.
- **Immersion, engagement and flow**.
- **Performance** by experimenting with alternative identities.
- **Identify, roles and embodiment**.
- **Intertextuality and integration** with other media forms (e.g., film, novels)
- **Narrative**, telling or generating story.
- **Geography**, the space in which the game takes place.

To address the issue of engagement for children, several of the measures identified above used interactive procedures and stimuli that would be more likely to foster interest and engagement (e.g., basketballs and hoops in the DANA task, comic-based format in the TEA-Ch 2). The literature review identified two papers that developed this further by specifically using a game-like paradigm.

Józsa, Barrett and Morgan (2017) developed a tablet-based task for children aged 3-8 years to assess school readiness and EF. The task is narrated by a cartoon bear who introduces each task as a game. EF is assessed through two tasks. The first task is titled the Picture Memory Game, in which the child is presented with an array of face down cards and is required to find matching pairs of cards, similar to the game ‘pairs’. The second task, named the ‘Size-Shape-Color Game’, asks children to sort stimuli according to colours and shapes into different baskets. The rules of how to categorise the stimuli (i.e., according to shape or colour) change periodically and the child is required to learn which rule is being used. The game-like element of this assessment tool appears relatively limited. Whilst the stimuli appear child-friendly and the instructions describe each task as a game, all stimuli presented are two-dimensional and there does not appear to be a story-line or overarching aim to motivate children to progress through each task. For example, when the child finishes the last task the bear says “goodbye” but no other reward is provided. In addition, there is no performance or identity aspect as the children do not adopt a character to play.

Johann and Karbach (2018) also created a computer-based assessment tool but embedded tasks in a ‘wizard’ storyline. The assessment tool is designed to measure EF in 8-11-year-olds and includes nine tasks to assess inhibition (Go/No-go, Flanker and Stroop-like), working memory (Complex span, Visuospatial span and n-Back) and flexibility (Task switching, Dual task and Continuous counting). The game element of this tool appears to be more developed than Józsa, Barrett and Morgan (2017). A storyline is presented and participants are invited into a ‘wizard kingdom’ in which they must complete tasks to defeat an ‘evil wizard’. After each task the child receives ‘magic power points’ and progressed through a map to the next task. Similarly to Józsa, Barrett and Morgan (2017), most of the tasks are two-dimensional and closely resemble standardised versions. This may limit engagement that could be facilitated by using a format that more closely resembles a game. Despite these limitations, both Józsa, Barrett and Morgan (2017) and Johann and Karbach (2018) found that the game-like format increased engagement.

More broadly than the computer format, Murphey (2017) used the commercial games Guess Who and Connect 4 as neuropsychological assessment measures for children with brain injuries and typically developing children. Murphey (2017) concluded the game-like measures demonstrated concurrent validity with existing measures. Pavitt (2017) created a game named ‘The Alien Game’ based on the 20-Questions task, and similar to the game Guess Who, to measure abstraction. Pavitt (2017) found participants rated the game-like measure as more enjoyable than existing measures and concluded that, with development, The Alien Game could be used as a formal cognitive test.

1.6 Summary of the Literature

The concept of EF is not well defined and numerous definitions have been used (Guare, 2014; Jurado & Rosselli, 2007), although there is agreement on the relevance of the frontal lobes (Diamond, 2013; Jurado & Rosselli, 2007; Miyake et al., 2000). Similarly, several models of EF have been proposed, however there is some consensus that Miyake et al. (2000) three-factor model is useful in both adults and children (Lehto et al., 2003). Although it was initially thought that EF did not develop until early adulthood, recent studies have identified that EF skills emerge in childhood. The development is not linear, with several periods of comparatively rapid development followed by a gradual maturation through late adolescence and early adulthood (Anderson, P., 2002; Anderson, V., 2002; Best & Miller, 2010, Goldstein & Naglieri, 2013).

Numerous tests and test batteries are used to measure EF in children, however these have mostly been developed originally for adults (Anderson, V., 2001). There are several challenges with the current assessment methods including: the definition of EF, developmental needs, specificity, ecological validity, cultural bias and engagement. In order to address some of these challenges, V. Anderson (2002) warns against the over-reliance of quantitative data and emphasises the need to utilise qualitative data (e.g., motivation, attention, strategy, emotional state) to aid the interpretation of EF test for children. Similarly, contextual information such as the child’s performance in school should be taken into account (Anderson, V., 2002; Fernández et al., 2014; Henry & Bettenay, 2010). To address the issue of engagement, researchers have utilised game-like paradigms and whilst both are limited by the two-dimensional format, they have been shown to be promising at increasing engagement (Johann & Karbach, 2018; Józsa, Barrett & Morgan, 2017).

1.7 Present Study

1.7.1 Aims and rationale

The present study aims to establish the feasibility of assessing executive functioning in children using novel, game-like, computerised tasks. Although computerised versions of standardised tests exist (e.g., the CANTAB), very few have utilised a game-like paradigm. A disadvantage of both Józsa, Barrett and Morgan (2017) and Johann and Karbach (2018) is that the tests are two-dimensional and closely resemble standardised versions.

This study therefore aims to develop a computerised novel measure, named Dragon Adventure, that is embedded within a three-dimensional space and appears more convincingly to replicate a game, rather than a neuropsychological test. Game-like paradigms have the potential to increase engagement and interest for children, therefore increasing the likelihood that the test can accurately capture the child's true ability. This may be of particular benefit to children where EF difficulties are suspected as a game-like paradigm could provide a more relaxed testing environment as compared to current standardised tests. In addition to the game-like procedure there are several potential benefits to using a computerised format. For clinicians perhaps the clearest benefit is in the reduction of administration and scoring errors (Goldstein & Naglieri, 2013), as this can be delivered by the program itself. Computerised formats also allow clinicians to streamline administration and there are opportunities to easily capture detailed data (e.g., precise time, accuracy, trail-by-trail performance and inter-trial interactions), which can in turn inform the formulation and intervention for the child. Finally, there are also opportunities to reduce cultural bias, created by test familiarity and language, by using the visual domain and creating a novel three-dimensional world in which the assessment takes place.

1.7.2 The novel measure: Dragon Adventure

Dragon Adventure was created using the software engine Unity 3D (Unity Technologies, 2018). It is a popular game engine used for developing three-dimensional games by both professional and novice developers. The Unity engine was chosen due to the extensive developer resources, rich graphics capabilities, free and low-cost gaming art assets, and the free licensing plan for projects of this size.

Dragon Adventure is comprised of three main tasks embedded within inter-connecting platforms and an overarching storyline. Several features were chosen to enhance the game-like protocol. Firstly, cartoon dragons were chosen as child-friendly main characters (Appendix A). Secondly, the player is able to move and play with the character in a large three-dimensional digital space that aims to maximise engagement (see Appendix B for examples). Lastly, a storyline was created in which the main character is lost in a 'Dragon Kingdom'. To return home the player must complete three tasks, at the end of each task the player is rewarded with a 'gem' which unlocks a gate to the next platform (Appendix C), until ultimately the dragon is able to return home (Appendix D).

Each of the three tasks created aim to measure the domains of inhibition, working memory and flexibility. In the first task, Dragon Dash, (Appendix E) the dragon runs along a moving platform and the player must avoid multiple obstacles, similar to 'Endless Runner' platform games. To assess inhibition, the task adopts a Stroop-like paradigm whereby the controls are either inversed (e.g., press left to move right) or not (e.g., press left to move left). The controls switch every 30 seconds and performance is measures by error rate. The second task, Dragon Sequence, (Appendix F) is adapted from the Spatial Spans task (Wechsler & Naglieri, 2006). The dragon is stuck behind a gate with nine squares on it. To assess working memory, a sequence of lights is demonstrated on the squares and the player must replicate the pattern in the reverse order. Performance can be measured in three ways: the longest correct sequence, total number correct, and number of errors per attempt. The final task, Dragon Hunt, (Appendix G) is adapted from the Trail Making Test (Reitan, 1992). The player must collect 10 blue eggs, 10 red eggs, 10 blue crystals, and 10 red crystals. To assess flexibility, the eggs and crystals can only be collected by alternating between the shape and colour (i.e., first the player collects the blue egg, then switch *shape* to blue crystal, switch *colour* to red crystal and so on). Performance is measured by error rate and completion time.

In order for Dragon Adventure to demonstrate utility as a neuropsychological test, it must demonstrate concurrent criterion validity with existing measures of executive function. As such, the following standardised measures were administered: D-KEFS Colour-Word Interference test, parts 1, 2 and 3 (Delis et al., 2001); D-KEFS Trail Making Test, parts 2 and 4 (Delis et al., 2001) and WNV Spatial Span, forwards and backwards (Wechsler & Naglieri, 2006).

Similarly, the measure should demonstrate acceptable levels of predictive criterion validity for it to be of use in making interpretations about a child's real-world level of functioning. To explore whether performance on Dragon Adventure correlated with real-world functioning the rating instrument Childhood Executive Functioning Inventory (CHEXI; Thorell & Nyberg, 2008) was given to a teacher who knew the child well.

Finally, to measure engagement, a visual analogue scale (Appendix H) was used at the end of each task. Participants were asked to rate how enjoyable they found the task on a scale of 1 – 5.

1.7.3 Research questions

The study aimed to address the following research questions:

1) Is Dragon Adventure more engaging for children than standardised measures of executive functioning?
2) Does performance on Dragon Adventure correlate with standardised measures of executive functioning (concurrent criterion validity)?
3) Does performance on Dragon Adventure correlate with a teacher-rated measure of executive functioning (predictive criterion validity)?

2. METHODS

2.1 Epistemology

Epistemology is concerned with the theory of knowledge (Ferrier, 1854). Several epistemological positions exist and are underpinned by different sets of assumptions about “what and how we can know” (Willig, 2012, pg. 10). It is important for researchers to reflect upon their epistemological position, as the research questions that shape the work are inevitably based on assumptions about the world and what can be known about it.

Willig (2012) categorises epistemological positions into three types of knowledge: realist, phenomenological, and social constructionist. Realist knowledge assumes that a ‘truth’ exists independently of a researcher’s awareness of it. Broadly speaking, realist approaches can be direct or critical. Direct realism reasons that the ‘truth’ can be measured directly (i.e., that the research data gathered directly represents reality). Critical realism assumes that there is an objective reality but that it cannot be directly observed (i.e., that the research data gathered is influenced by the context in which it was generated). Phenomenological approaches do not attempt to uncover objective ‘truth’, instead they aim to understand the research participant’s subjective experience. Finally, social constructionist approaches assume that knowledge about the world is generated and constructed through language and social discourses.

This thesis reflects a critical realist position. This position allows the researcher to attempt to explore and measure constructs (e.g., ‘executive functioning’), that are assumed to exist in the world independently of the researcher. However, the approach also allows the researcher to acknowledge that the term ‘executive functioning’ has been generated by previous attempts to measure the concept. Therefore, the term ‘executive functioning’ now acts as a vehicle through which the research is constructed, and observations are made. Nevertheless, it is hoped that the research can contribute to the understanding of how to measure ‘executive functioning’.

It should be noted that the researcher believes that, whilst some underlying cognitive processes may exist independently, the concept of ‘executive functioning’ is socially constructed. What is considered acceptable or ‘normal’ is inevitably influenced by the social context. For example, deficits in ‘executive function’ are often associated with the diagnosis of ADHD (Goldsteing & Naglieri, 2013). This has been strongly criticised for being a Westernised understanding of behaviour that has, in different places and times, been considered to be on the spectrum of ‘normal’ behaviour, in particular pathologizing the behaviour of boys (Timimi & Timimi, 2015). As such, the research data should be interpreted with caution and situated within the social context it has arisen from.

2.2 Design

The study used a mixed methods design. A cross-sectional correlational design was used to investigate the relationships between the novel game-like measures with established measures and teacher-rated measures of executive function. The performance of ‘typically developing’ children on Dragon Adventure was compared to the scaled scores obtained on the D-KEFS Colour-Word Interference test (Delis et al., 2001), D-KEFS Trail Making Test (Delis et al., 2001), WNV Spatial Span (Wechsler & Naglieri, 2006), and the CHEXI (Thorell & Nyberg, 2008). In addition, a within-subjects comparison of visual-analogue ratings was conducted to explore whether there were differences between the enjoyment ratings of each task.

2.3 Recruitment, Inclusion and Exclusion Criteria

Children were recruited from a secondary school in London. The inclusion criteria were the following: children aged 11-12 years who were able to understand and read information in English. A limited age range was chosen as creating an age-appropriate measure across childhood was beyond the scope of the study and not yet clinically indicated. This age range was chosen as it is thought that EF is relatively well established by 11 and 12 years (Best & Miller, 2010).

The exclusion criteria were the following: children who had a diagnosis of a learning disability or those who did not speak English. Within the school, children with learning or language needs are placed in the same form group. As such, children from this form group were excluded from the study.

2.4 Dragon Adventure: Test Development

Dragon Adventure consists of three subtests adapted from established formats. It was created using the software engine Unity 3D (Unity Technologies, 2018) and a range of art ‘assets’ downloaded from the Unity Asset Store (Unity Technologies, 2020). See Appendix I for a list of artwork and acknowledgements to the authors.

All written instructions were provided on-screen, in large rectangular text boxes, in the centre of the screen (Appendix J). Participants were presented with a short sentence, or part-sentence, before clicking the ‘continue’ button to receive the next piece of information. Once they reached the end, the text box disappeared off-screen so that navigation may continue. To navigate the main character the participants used the left, right, up and down arrow keys as well as the space bar. The laptop trackpad was used to click buttons on-screen. See section 1.7.2 for an overview of Dragon Adventure and game-like features. Screenshots of Dragon Adventure can be found in Appendix A – G.

2.4.1 Dragon Dash

This subtest was designed to measure inhibition. The cartoon dragon character runs along a moving ground platform. The platform contains multiple obstacles, the aim is to avoid the obstacles and reach the end of the platform to collect a hidden ‘gem’.

The participants were first presented with the instructions (Appendix K), then began the task. Four types of obstacle were presented in a random (seed generated) order, so each participant was presented with the same sequence of obstacles. The obstacles were as follows: left-hand block (Appendix L), right-hand block (Appendix M), jumping block (Appendix N) and ducking block (Appendix O). Each obstacle was presented on a separate platform (Appendix P), the platforms were generated in front of each-other so that the main character could run smoothly across (Appendix E).

The participant used the left, right, up, and down arrow keys to avoid the obstacles. To measure inhibition, a Stoop-like paradigm was utilised. The arrow keys began by being inverted (i.e., left moves right, right moves left, up moves down [duck] and down moves up [jump]). After 30 seconds the participant was presented with an audible beep to indicate that the controls had switched (i.e., left moves left, right moves right, up moves up [jump] and down moves down [duck]). Every 30 seconds the beep sounded to indicate that the controls had been reversed again, thus switching between congruent and incongruent trials.

Platforms were generated for 5 minutes exactly, once the participant had navigated the final platform they were taken to a ‘celebration scene’ (Appendix Q) where they received the reward ‘gem’. An error was recorded each time they hit a novel obstacle; the background timer continued. Performance was measured by the total number of errors.

2.4.2 Dragon Sequence

This subtest was designed to measure working memory and was adapted from the WNV Spatial Spans task (Wechsler & Naglieri, 2006). The main character is ‘stuck’ behind a large gate. The participant was presented with a sequence of lights which they needed to replicate backwards in order to succeed, unlock and open the door to collect a ‘gem’.

The participants were first presented with the instructions (Appendix R), then began the task. Nine squares were presented on the gate (Appendix F), the participant was presented with a sequence of lights, indicated by the relevant squares turning red. The square was lit red for one second exactly then returned to grey, the next square was then lit. The number of lights in the sequence gradually increased, from two to a maximum of nine. Sequences were generated pseudo-randomly, where the same number could not occur more than once. The sequences presented are given in Table 4.

Once the sequence was complete, the participant was required to replicate the sequence, in the reverse order. Participants used the laptop trackpad to click the corresponding squares. If the participant clicked an incorrect square, it was recorded as an error, and the next sequence was presented. If the participant made two errors consecutively, the subtest concluded, and the participant collected a ‘gem’. Performance can be measured in three ways: the longest correct sequence, total number correct, and number of errors per attempt.

Table 4. Sequences presented to participants.

Sequence number	Sequence order
1	2, 8
2	6, 7
3	8, 1, 3
4	4, 1, 8
5	5, 7, 8, 3
6	6, 8, 4, 7
7	9, 7, 1, 2, 4
8	8, 5, 2, 7, 6
9	1, 6, 7, 8, 4, 9
10	3, 4, 9, 8, 6, 7
11	4, 2, 7, 3, 8, 6, 5
12	7, 3, 6, 4, 1, 9, 5
13	2, 7, 5, 1, 3, 4, 6, 8
14	2, 4, 7, 5, 8, 3, 6, 1
15	9, 5, 1, 6, 2, 3, 4, 7, 8
16	2, 8, 9, 3, 7, 5, 6, 4, 1

2.4.3 Dragon Hunt

This subtest was designed to measure flexibility and was adapted from the Trail Making Test (Reitan, 1992). The main character is presented with an open landscape. The participant must collect ‘eggs’ (Appendix S) and ‘crystals’ (Appendix T), by alternating between colour and shape, to succeed and receive a ‘gem’.

The participants were first presented with the instructions (Appendix U), then began the task. Four types of stimuli were presented: 10 blue eggs, 10 red eggs, 10 blue crystals and 10 red crystals. In the top left-hand corner of the screen a counter showed the total number of 'eggs' and 'crystals' left to collect. Participants were only able to collect the stimuli in a specific order, determined by alternating the colour and the shape. Participants were instructed to begin by collecting a 'blue egg'. They then needed to switch *shape*, to collect a 'blue crystal', then *colour* to 'red crystal', then *shape* to 'red egg' and so on. Performance was measured by total completion time and total number of errors.

2.5 Existing Measures

In addition to the novel Dragon Adventure measure, several established measures were used to explore the concurrent and predictive criterion validity of the novel game-like measure.

2.5.1 D-KEFS Colour-Word Interference test

The D-KEFS Colour-Word Interference test (CWIT; Delis et al., 2001) parts one, two and three were used. This is based on the classic Stroop paradigm and is designed to measure inhibition in the verbal domain. It includes four conditions: Colour Naming, Word Naming, Inhibition and Inhibition/Switching. In condition one the participant is required to name patches of colour (e.g. ■, ■, ■). In condition two the participant is required to name printed written words (e.g., red, blue, green). In condition three the participant is required to name the ink colour a word is printed in and inhibit word reading (e.g., blue, green, red). Performance is measured by completion time and errors. Contrast sores allow the examiner to distinguish between inhibition/ switching skills, simple processing speed and and lower-level abilities.

2.5.2 WNV Spatial Span

The WNV Spatial Span (SS; Wechsler & Naglieri, 2006), forwards and backwards trials were used. It is adapted from the WISC-IV Integrated (Wechsler et al., 2004) span tasks and is designed to measure working memory. It has a forwards and backwards condition. In the forwards condition, the examiner points to a sequence of boxes on a board and the participant is required to repeat the sequence in the same order. In the backwards condition, the examiner points to a sequence of boxes on the same board and the participant is required to repeat the sequence in the reverse order. An illustration of the task is provided in Figure 2. The task concludes when the participant makes two consecutive errors. Performance is measured by the total correct, longest correct sequence and combined forwards and backwards scores.

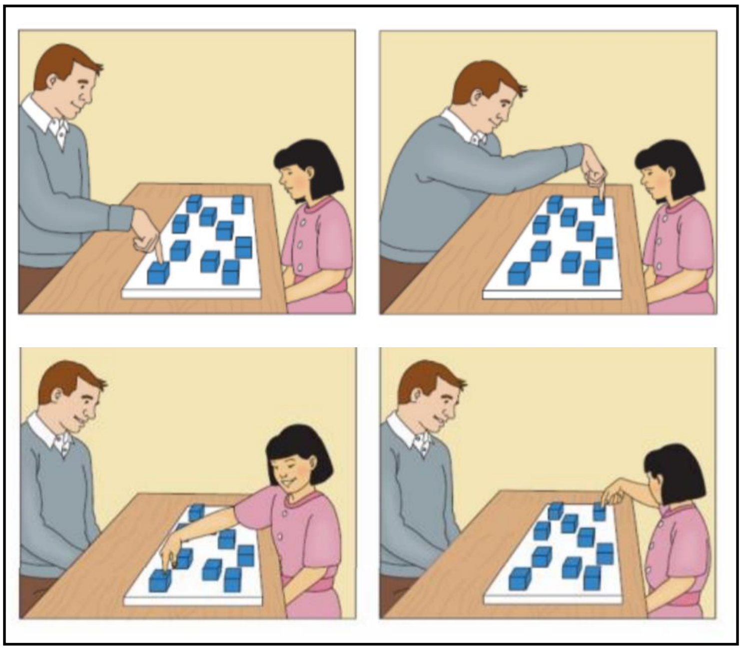


Figure 2. Illustration of WNV Spatial Span task (Naglieri, 2010).

2.5.3 D-KEFS Trail Making Test

The D-KEFS Trail Making Test (TMT; Delis et al., 2001) parts two and four were used. It is developed from the Reitan (1992) Trail Making Test and is designed to measure flexibility in the visual domain. It includes five conditions in total: Visual Scanning, Number Sequencing, Letter Sequencing, Number-Letter Sequencing, and Motor Speed. These allow the examiner to attempt to distinguish the different skills used in the task. In part two, Number Sequencing, the participant is required to connect numbers in numerical order (e.g., 1-2-3). In part four, Number-Letter Sequencing, the participant is required to connect numbers in numerical order, and letters in alphabetical order, by alternating between numbers and letters (e.g., 1-A-2-B). Performance is measured by completion time and errors. Contrast scores are derived by subtracting the completion time scaled score for condition component conditions.

2.6 Childhood Executive Functioning Inventory (CHEXI)

The Childhood Executive Functioning Inventory (CHEXI; Thorrell & Nyber, 2008) was used to assess criterion predictive validity. The CHEXI is a 24-item questionnaire intended to be used by parents and teachers. It includes four subscales measuring: working memory (9 items), planning (4 items), inhibition (6 items), and regulation (5 items). Factor analysis in children aged 8-11 years in Belgium and Sweden (Catale, Meulemans & Thorell, 2015) identified two factors, working memory (working memory and planning subscales) and inhibition (inhibition and regulation subscales). Normative data has not yet been gathered, however Catale et al. (2015) administered the CHEXI to 242 typically developing children. The means, standard deviations and proposed cut-off scores are detailed in Table 5.

Table 5. Combined means, standard deviations and proposed cut-off scores for the CHEXI reported by Catale et al. (2015, pg. 5).

	Mean (SD)	Cut-off
Working Memory	24.05 (7.89)	34-35
Inhibition	25.65 (7.89)	32-34

2.7 Visual analogue scale

A 5-point Likert visual analogue scale was used to measure the acceptability of the novel game-like and established measures (Appendix H). At the end of each of the three subtests of Dragon Adventure, participants were asked to rate how enjoyable the task was. They also rated their enjoyment of the D-KEFS Colour-Word Interference test, the WNV Spatial Span, and the D-KEFS Trail Making Test.

2.8 Ethics

Ethical approval was gained from the University of East London Research Ethics Committee (Appendix V). The study was conducted within a secondary school in London. The headteacher of the school was approached and asked whether they would allow the school to participate in the study. The headteacher provided permission to recruit year 7 students. Consent was firstly sought from parents and/or guardians. Information and consent forms were distributed to parents and/or guardians (Appendix W & X). Those who consented to their child participating, returned a signed consent form via the form tutor. All children for whom parental consent was provided met the inclusion criteria for the study. The researcher then met with each child and provided the child information and consent form (Appendix Y). The children had the opportunity to discuss the study with the researcher and ask questions. Once it was clear the child understood the study, the researcher sought consent. All children agreed to participate and signed the child consent form.

Parents and/or guardians and children were fully informed regarding the study and no deception was used. No significant risks to participants physical, emotional or psychological wellbeing were identified. Nevertheless, all participants were monitored for signs of distress and/or fatigue and were advised that they could take breaks if they wished. No participant became distressed during the study. Parents and/or guardians and children were advised that they could withdraw the data at any point without having to provide a reason and without disadvantage to themselves.

Each child was allocated a unique, anonymising identification code which was stored in a separate, password-protected database against participant names. This was kept so that participant data could be withdrawn if requested. Identification codes were used for all analysis databases. All paper consent and scoring forms were scanned and stored in a password-protected file on the UEL H: Drive. All paper consent forms and score sheets were destroyed after scanning.

2.9 Materials and Procedure

The study used:

- The novel computer measures (Appendix A – G) and laptop.
- D-KEFS Colour-Word Interference test, parts 1, 2 and 3 and scoring sheets (Delis et al, 2001).
- WNV Spatial Span, forwards and backwards and scoring sheets (Wechsler & Naglieri, 2006).
- D-KEFS Trail Making Test, parts 2 and 4, and scoring sheets (Delis et al, 2001).
- The Childhood Executive Functioning Inventory (CHEXI) (Thorell & Nyberg, 2008).
- Information sheets (Appendix W & Y)
- Consent forms (Appendix X & Y)
- Visual analogue scale (Appendix H)
- Pen and paper
- Timer

Consent from parents/guardians was collected prior to participants completing the study. Before starting the tasks, the children were informed about the study and consent was requested. All children provided consent in addition to parent/guardian consent.

Testing took place in a quiet room in the school, the participants sat opposite the researcher at a table. The participants were first presented with Dragon Adventure. This was presented on a laptop approximately 20cm from the participant. All the instructions appeared on the screen, but participants were also advised that they could ask the researcher if they did not understand. Some participants did require additional instructions, in this case the written instructions were repeated verbatim. The implications are discussed in the discussion section. Participants progressed through each of the three tasks on the laptop until completion. At the end of each of the three tasks the game play was paused, and participants were asked to rate how enjoyable each task was, using the visual analogue scale.

Participants were then presented with the D-KEFS Color Word Interference Test (Delis et al., 2001). The subtests Colour-naming, Word Reading and Inhibition were administered in this order. The WNV Spatial Spans (Wechsler & Naglieri, 2006), forwards and backwards, was then administered. Finally, the D-KEFS Trail Making Test (Delis et al., 2001), parts 2 and 4, were administered. After the participants had completed all subtests of each measure, they were asked to rate how enjoyable they found the task using the visual analogue scale. The participants were then informed that the study was complete and offered the opportunity to ask questions. Separately, a teacher who knew each participant well was approached and asked to complete the CHEXI.

2.10 Participants

<u>2.10.1 Demographic data</u>

There were 21 participants, 11 males and 10 females, ranging in age from 135 to 147 months (Mean = 11.70 years, SD = 0.36). There were 16 primary English language speakers (PEL) and 5 participants with English as an additional language (EAL); a Chi-square test indicated that this difference was significant, X^2 (df, 1) = 5.76, p = .027. This group size should be taken into account when examining the effect of PEL or EAL groups.

There were no significant differences in age between sex ($U[N = 11, 10] = 54.00$, z = -0.07, p = .973) nor English language groups ($U[N = 5, 16] = 28.00$, z = -0.99, p = .354).

2.10.2 Sample Characteristics

Tables 6-8 give the descriptive statistics for the participants' scores on the D-KEFS Colour-Word Interference test, WMV Spatial Span and D-KEFS Trail Making Test. Raw scores have been converted to age-scaled scores, and are provided for sex and language groups, as well as the overall sample. The scores obtained in this sample were compared to normative data (Mean = 10, SD = 3) to indicate how typical the sample children were compared to children of the same age. In the overall sample, scores for the CWIT and SS were slightly above the norms. Scores for the TMT were slightly below normative data. Male participants scored slightly higher across most measures and participants with EPL scored somewhat higher on most measures than participants with EAL. However, Mann-Whitney U tests found that none of these differences were reliable (all $p > .073$).

Table 6. Participants' scaled scores on D-KEFS Colour-Word Interference test, mean (standard deviation).

		N	Combined CN and WR	Inhibition Time	Inhibition Errors
Sex	Male	11	11.45 (1.92)	12.00 (2.57)	11.09 (4.74)
	Female	10	10.60 (2.01)	10.20 (3.36)	9.80 (2.74)
Language	EPL	16	11.31 (1.82)	11.81 (2.14)	11.13 (4.16)
	EAL	5	10.20 (2.39)	9.00 (4.64)	8.40 (1.82)
Overall sample		21	11.05 (1.96)	11.14 (3.04)	10.48 (3.88)

Table 7. Participants' scaled scores on WMV Spatial Span, mean (standard deviation).

		N	Forwards	Backwards	Combined
Sex	Male	11	10.27 (2.53)	11.91 (2.17)	11.55 (1.75)
	Female	10	10.30 (2.58)	10.00 (2.82)	10.30 (2.95)
Language	EPL	16	10.25 (2.30)	11.31 (2.47)	11.19 (2.23)
	EAL	5	10.40 (3.36)	10.00 (3.16)	10.20 (3.11)
Overall sample		21	10.29 (2.49)	11.00 (2.93)	10.95 (2.42)

Table 8. Participants' scaled scores on D-KEFS Trail Making Test, mean (standard deviation).

		N	Number Sequencing Time	N-L Sequencing Time
Sex	Male	11	9.73 (2.37)	9.18 (3.06)
	Female	10	8.70 (2.87)	9.60 (9.95)
Language	EPL	16	8.94 (2.46)	9.94 (2.49)
	EAL	5	10.20 (3.11)	7.60 (3.85)
Overall sample		21	9.24 (2.61)	9.38 (2.94)

Participants' scores on the CHEXI were converted age-scaled scores using the means from Catale et al. (2015), Table 9 displays the descriptive statistics. Scores for the overall sample were in line with normative data for the Working Memory subscale and somewhat better for the Inhibition subscale. Males achieved somewhat better scores compared to female participants on both Working Memory and Inhibition subscales. Participants with EAL scored slightly poorer on both subscales than participants with EPL. Again, Mann-Whitney U tests found that none of these differences were reliable (all p > .080).

Table 9. Participants' scaled scores on the Childhood Executive Functioning Inventory (CHEXI) Working Memory and Inhibition subscales, by sex, language and overall sample, mean (standard deviation).

		N	Working Memory	Inhibition
Sex	Male	11	11.18 (2.96)	13.00 (3.03)
	Female	10	8.20 (4.42)	9.90 (4.56)
Language	EPL	16	10.50 (3.83)	12.19 (3.73)
	EAL	5	7.40 (3.65)	9.40 (4.78)
Overall sample		21	9.76 (3.94)	11.52 (4.06)

3. RESULTS

3.1 Methods of Analysis

The data was analysed using SPSS (Statistical Package for the Social Sciences) Version 26.0. Boxplots were generated to check for outliers and data entry errors were corrected. The D-KEFS CWIT, WNV Spatial Spans and D-KEFS Trail Making Tests were scored according to the test instructions and scaled scores generated. The CHEXI was scored according to the instructions. Preliminary ‘normative’ data was derived from the Catale et al. (2015) study and scaled scores were calculated.

Descriptive statistics for the cognitive data are given in Table 10. Histograms, skeweness (>1), kurtoisis (>3) and Shapiro-Wilk’s test were used to examine normality of distributions. Statistics that indicated non-normality are identified in bold in Table 10 and are described below:

- *Dragon Sequence – Errors Per Attempt, Dragon Hunt – Total Errors, and D-KEFS TMT – N-L-S Errors* all appear positively skewed. Scores are clustered around a ‘low’ value, with a long ‘tail’ of higher scores.
- *Dragon Hunt – Time* also appears positively skewed and somewhat kurtotic, with most participants completing the measure between 200-300 seconds and a flat distribution of remaining time values.
- *D-KEFS CWIT – Inhibition Time, WNV Spatial Spans – Backwards Longest Correct, and CHEXI – Inhibition* all appear negatively skewed. Scores are clustered around a ‘high’ value, with a long ‘tail’ of lower scores.
- *CHEXI – Working Memory* was found to be reliably ‘non-normal’ by the Shapiro-Wilk test. Inspection of a histogram suggests a bi-model distribution, with most participants scoring either 8-9 or 14. However, due to the small sample size, the distribution pattern is difficult to assess.

As 50% of the variables were indicated to be non-normally distributed, the sample size is small (N = 21), and as the English language groups are different in size, non-parametric tests were used for all further analysis.

Mann-Whitney tests were used to compare between-group means for sex and language groups. Spearman's rank correlations were used to assess the relationships within the three novel measures and between the novel and existing measures. Wilcoxon's Signed-Rank tests were used to examine differences in acceptability ratings on the visual analogue scale. Effect sizes were used to interpret results, specifically Cohen's (1988) recommendations were used to interpret the size of effects: small, .10-.29; moderate, .30-.49; and large, ≥.50. For moderate and large effect sizes significance values are also reported, although it should be noted that some are non-significant ($p > 0.05$) and should thus be interpreted with caution. Finally, Cronbach's alpha and Spearman-Brown formula were used to investigate the internal consistency of Dragon Adventure.

Table 10. Descriptive statistics for cognitive data

		Mean	SD	Min.	Max.	Skewness z-score	Kurtosis z-score	Shapiro-Wilk Sig (2-sided)
Novel Measures								
Dragon Dash	Total Errors	22.90	7.91	10	38	0.377	-0.557	.482
Dragon Sequence	Total Correct	5.52	3.14	0	11	-0.100	-0.938	.650
	Longest Correct	4.33	1.85	0	7	-0.386	-0.070	.362
	Errors Per Attempt	0.41	0.22	0.15	1.00	**1.379**	1.383	**.003**
Dragon Hunt	Time (m)	6.66	3.49	3.03	17.05	**1.48**	2.46	**.003**
	Total Errors	10.86	9.71	0	38	**1.263**	1.464	**.011**
Existing measures								
D-KEFS CWIT	Inhibition Time (SS)	11.14	3.04	4	15	**-1.167**	0.657	**.011**
	Inhibition Errors (SS)	9.95	2.66	4	14	-0.635	-0.291	.298
WNV Spatial Spans	Backwards Total Correct (SS)	11.00	2.63	5	15	-0.201	-0.095	.405
	Backwards Longest Correct	5.24	0.94	3	6	-0.921	-0.254	**>.001**
	Combined (SS)	10.95	2.42	5	16	-0.197	1.228	.460
D-KEFS Trail Making Test	N-L-S Time (SS)	9.38	2.94	3	13	-0.754	-0.299	.059
	N-L-S Errors	1.29	1.23	0	4	0.639	-0.552	**.011**
	N-S and N-L-S Contrast (SS)	10.14	3.53	3	16	-0.638	0.129	.192
CHEXI	Working Memory (SS)	9.76	3.94	1	14	-0.837	0.463	**.008**
	Inhibition (SS)	11.52	4.06	1	16	**-1.047**	0.731	**.021**

3.2 Performance on the Novel Measures

3.2.1 Sex

A summary of the scores on Dragon Adventure by sex groups, is provided in Table 11. Male participants scored slighter better on all variables apart from *Dragon Hunt – Total Errors*, where they made more errors than female participants. Mann-Whitney tests indicate that these differences are reliable for *Dragon Sequence – Total Correct* and *Longest Correct*. For *Total Correct*, males achieved significantly more correct trials (*Mdn* = 7.00) compared to females (*Mdn* = 4.00), U = 24.50, z = -2.16, p = .029, r = -0.47. Similarly, for *Longest Correct*, males were able to recall significantly longer sequences (*Mdn* = 5.00) than females (*Mdn* = 3.50), U = 21.50, z = -2.39, p = .016, r = -0.52. It should be noted, however, that these variables are closely related. Mann-Whitney tests indicated all other comparisons were non-significant (all p > .114).

Table 11. Participant's scores by sex on Dragon Dash, Dragon Sequence and Dragon Hunt measures, mean (standard deviation).

		Male			Female		
		N	Mean	SD	N	Mean	SD
Dragon Dash	Total Errors	11	21.27	7.06	10	24.70	8.76
Dragon Sequence	Total Correct	11	6.91	3.24	10	4.00	2.31
	Longest Correct	11	5.09	2.07	10	3.50	1.18
	Errors Per Attempt	11	0.37	0.25	10	0.45	0.18
Dragon Hunt	Time (m)	11	5.74	2.67	10	7.68	4.11
	Total Errors	11	11.36	12.24	10	10.30	6.52

3.2.2 Language

A summary of the scores on Dragon Adventure by language groups, is provided in Table 12. The EPL group scored somewhat better on all variables apart from *Dragon Hunt – Total Errors*, where they made more errors than EAL group. Mann-Whitney tests indicate that these differences are reliable for *Dragon Dash – Total Errors* and *Dragon Sequence – Longest Correct*. For *Dragon Dash – Total Errors*, the EPL group made significantly fewer errors (*Mdn* = 19.50) compared to the EAL group (*Mdn* = 28.00), U = 12.00, z = -2.31, p = .019, r = -0.50. For *Dragon Sequence – Longest Correct*, the EPL group were able to recall significantly longer sequences (*Mdn* = 5.00) than the EAL group (*Mdn* = 3.00), U = 65.50, z = 2.13, p = .032, r = 0.46. Mann-Whitney tests indicated all other comparisons were non-significant (all p > .05).

Table 12. Participants' scores by language group on Dragon Dash, Dragon Sequence and Dragon Hunt measures, mean (standard deviation).

		EPL			EAL		
		N	Mean	SD	N	Mean	SD
Dragon Dash	Total Errors	16	20.81	7.20	5	29.60	6.66
Dragon Sequence	Total Correct	16	6.25	3.02	5	3.20	2.49
	Longest Correct	16	4.75	1.84	5	3.00	1.23
	Errors Per Attempt	16	0.38	0.21	5	0.51	0.24
Dragon Hunt	Time (m)	16	5.91	2.41	5	9.08	5.41
	Total Errors	16	11.25	10.92	5	9.60	4.67

3.2.3 Age

A summary of Spearman's rank correlation coefficients between age and the three novel measures can be seen in Table 13. Very small correlations can be seen between age and *Dragon Dash – Total Errors, Dragon Sequence – Total Correct, Dragon Sequence – Longest Correct,* and *Dragon Hunt – Total Errors.* There is almost no correlation between age and *Dragon Sequence – Errors Per Attempt* and *Dragon Hunt – Time*. None of these correlations were found to be reliable (all r <.333)

Table 13. Spearman's rank correlations between age and novel measures.

		Age	Dragon Dash – Total Errors	Dragon Sequence – Total Correct	Dragon Sequence – Longest Correct	Dragon Sequence – Errors Per Attempt	Dragon Hunt - Time	Dragon Hunt – Total Errors
Age	Coefficient	1.000	.254	-.263	-.284	.012	.061	.212
	Sig.		.266	.250	.212	.957	.793	.356

3.2.4 Motor speed

A summary of Spearman's rank correlation coefficients between TMT number sequencing (NS) task and the novel measures can be seen in Table 14. The TMT NS task was used as an approximation of motor-speed. Very small correlations can be seen between *TMT NS Time* and *Dragon Hunt – Time* and *Dragon Hunt – Total Errors*. There is almost no correlation between the *TMT NS Time* and remaining novel measures. None of these correlations were found to be reliable (all r >.333).

Table 14. Spearman's rank correlations between TMT number sequencing time and novel measures.

		TMT NS Time	Dragon Dash – Total Errors	Dragon Sequence – Total Correct	Dragon Sequence – Longest Correct	Dragon Sequence – Errors Per Attempt	Dragon Hunt – Time	Dragon Hunt – Total Errors
TMT NS Time	Coefficient	1.000	-.026	-.007	-.037	-.105	-.258	-.230
	Sig.		.911	.975	.873	.651	.260	.316

3.3 Associations within and between Dragon Adventure

A summary of Spearman's rank correlation coefficients, between and within the three novel measures, can be seen in Table 14. Moderate and large effect sizes are marked in bold.

When comparing variables within the same measure, there was a large correlation between *Dragon Sequence - Total Correct* and *Dragon Sequence - Longest Correct* (r = .978, *p* <.001). This correlation is so large that, in this sample, these two variables could be considered to be measuring the same thing. Accordingly, there was also a large negative correlation between *Dragon Sequence – Errors Per Attempt* and both *Dragon Sequence – Total Correct* (r = -.802, *p* <.001) and *Dragon Sequence – Longest Correct* (r = -.732, *p* <.001). Thus, as participants' total and longest correct score improved, fewer errors per attempt were made.

For the *Dragon Hunt* task, there was a large correlation between *Time* and *Total Errors* (r = .784, *p* <.001). As participants' completion time decreased, fewer errors were made.

There is a large correlation between *Dragon Dash - Total Errors* and *Dragon Hunt Time* (r = .552, *p* = .009). As the number of errors on *Dragon Dash* goes up, so does the amount of time needed to complete the *Dragon Hunt* task. Similarly, there is also a positive moderate correlation with *Dragon Hunt – Total Errors* (r = .482, *p* = .053). There is a small to moderate correlation with *Dragon Sequence – Total Correct* and *Longest Correct* and no correlation with *Dragon Sequence – Errors Per Attempt.*

There appeared to be little correlation between *Dragon Sequence* and *Dragon Hunt* measures, with all correlations being small. The largest of these small correlations was between *Dragon Sequence – Errors Per Attempt* and *Dragon Hunt – Time* (r = .283, *p* = .214).

As *Dragon Sequence – Total Correct* and *Dragon Sequence – Longest Correct* are very highly correlated, only *Dragon Sequence – Total Correct* is used for all further analysis.

Table 15. Spearman's rank correlation coefficients within novel measures.

	Dragon Dash – Total Errors	Dragon Sequence – Total Correct	Dragon Sequence – Longest Correct	Dragon Sequence – Errors Per Attempt	Dragon Hunt - Time	Dragon Hunt – Total Errors
Dragon Dash Total Errors	1.000					
Dragon Sequence – Total Correct	-.279	1.000				
Dragon Sequence Longest Correct	**-.327**	**.978****	1.000			
Dragon Sequence Errors Per Attempt	.070	**-.802****	**-.732****	1.000		
Dragon Hunt - Time	**.552****	-.198	-.204	.283	1.000	
Dragon Hunt – Total Errors	**.428**	.103	.094	-.130	**.748****	1.000

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

3.4 Internal Consistency

As part of test development, reliability analyses were carried out on the three novel measures to assess their consistency. For the *Dragon Dash* test, a split-half reliability analysis was conducted by comparing the total errors made in the first and second halves of the test, as determined by time. The Spearman-Brown coefficient showed Dragon Dash to have acceptable reliability, $\rho = .751$.

For the *Backwards Span* test, Cronbach's alpha showed the measure to have good reliability, $\alpha = .858$. Deletion of the first trial indicated a small increase in reliability to $\alpha = .862$. Therefore, the test may be improved by considering the first trial as a 'practice trial' and disregarding the result.

For the *Dragon Hunt* test, a split-half reliability analysis was conducted by comparing the total errors made in the first and second halves of the test. This was determined by the point at which the participant had successfully collected half of the items. The Spearman-Brown coefficient showed the *Dragon Hunt* test to have good reliability, $\rho = .846$.

3.5 Associations with Established Performance-Based Measures

A summary of Spearman's rank correlation coefficients, between the novel measures and established performance-based measures, can be seen in Table 15. Moderate and large effect sizes are marked in bold.

For *Dragon Dash – Total Errors*, there was a moderate negative correlation with *CWIT – Inhibition Time* ($r = -.399$, $p = .073$) but no association with *CWIT – Inhibition Errors* ($r = -.190$, $p = .410$). This suggests that as errors on *Dragon Dash* increase, performance on the *CWIT – Inhibition Time* decreases. *Dragon Dash – Total Errors* is also negatively associated with *TMT – N-L-S Time* ($r = -.305$, $p = .179$), suggesting participants who made more errors on the novel *Dragon Dash* measure performed worse on *TMT – N-L-S* than participants who made fewer errors.

Dragon Sequence – Total Correct was associated with several existing measures. As expected, it was most strongly correlated with *Spatial Spans*. There was a strong positive correlation with *SS – Backwards Total Correct* (r = .561, *p* = .008) and *SS – Combined* (r = .501, *p* = .021), as well as a moderate positive correlation with *SS - Backwards Longest Correct* (r = .322, *p* = .154). There was also a moderate positive correlation with *CWIT – Inhibition Time* (r = .470, *p* = .032). *Dragon Sequence – Total Error Per Attempt* was strongly negatively correlated with *CWIT – Inhibition Time* (r = -.509, *p* = .018) and moderately negatively correlated with *CWIT – Inhibition Errors* (r = -.315, *p* = .164). This suggests that as errors increased on the novel *Dragon Sequence* measure, performance on the *CWIT* decreased.

Dragon Hunt – Time is strongly negatively associated with *CWIT – Inhibition Time* (r = -.787, *p* < .001). *Dragon Hunt – Total Errors* was also strongly negatively correlated with *CWIT – Time* (r = -.602, *p* = .004). This suggests as time and error rate improve on the novel *Dragon Hunt* task, so does completion time on the *CWIT – Inhibition* task. *Dragon Hunt – Time* was moderately negatively associated with *SS – Backwards Longest Correct* (r = -.420, *p* = .058), as was *Dragon Hunt – Total Errors* (r = -.348, *p* = .122). This suggests that as performance on the novel *Dragon Hunt* task increases, so does performance on *SS – Backwards*. Finally, there was a moderate negative correlation between *Dragon Hunt – Total Errors* and *TMT N-L-S Time* (r = -.342, *p* = .129). This suggests that as the errors on the novel *Dragon Hunt* task increases, performance on *TMT - N-L-S Time* decreases.

Table 16. Spearman's rank correlation coefficients between novel measures and existing performance-based measures.

	Dragon Dash - Total Errors	Backward spans - Total Correct	Backward spans - Errors/ attempt	Dragon Hunt - Time taken	Flexibility - Total errors
Colour-word: inhibition time	**-.399**	**.470***	**-.509***	**-.787****	**-.602****
Colour-word: inhibition errors	-.190	.276	**-.315**	-.159	.096
Spatial span: backward total	.046	**.561****	-.297	-.256	-.210
Spatial span: backward length	.020	**.322**	-.189	**-.420**	**-.348**
Spatial span combined	-.032	**.501***	**-.326**	-.259	-.184
TMT: combined time	**-.305**	.284	-.008	-.275	**-.342**
TMT: combined errors	-.143	.057	.100	.272	.200
TMT: contrast	-.252	.119	.117	-.104	-.177

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

3.6 Associations with the CHEXI

A summary of Spearman's rank correlation coefficients between the novel measures and the *CHEXI* can be seen in Table 16. For comparison, Table 17 displays the associations between the established measures and the CHEXI. Moderate and large effect sizes are marked in bold. The *CHEXI* was most strongly associated with *Dragon Sequence – Total Correct* on both the *Working Memory (*r = .510, *p* = .018*)* *and Inhibition* (r = .476, *p* = .029) subsets. This suggests that as performance on the novel *Dragon Sequence* measure improves, so do the ratings on the *CHEXI*.

There was also a small negative association between *Dragon Dash – Total Errors* and both subsets of the *CHEXI*, suggesting that as errors on *Dragon Dash* increased, EF skills were rated as worse on the CHEXI*. CHEXI Inhibition* was also weakly negatively associated with both *Dragon Hunt – Time* and *Dragon Hunt – Total Errors*. This suggests that as performance on the novel *Dragon Hunt* task improves, so does the ratings on the *CHEXI*. Overall, associations between the established measures and the *CHEXI* were similar to the associations found between the novel measures and the *CHEXI*.

Table 17. Spearman's rank correlation coefficients between the novel measures and the CHEXI.

	Dragon Dash – Total Errors	Dragon Sequence – Total Correct	Dragon Sequence – Errors Per Attempt	Dragon Hunt - Time	Dragon Hunt – Total Errors
CHEXI Working Memory	-.242	**.510***	-.195	-.157	-.158
CHEXI Inhibition	-.262	**.476***	-.197	-.202	-.222

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table 18. Spearman's rank correlation coefficients between the established measures and the CHEXI.

	Colour-word: inhibition time	Colour-word: Inhibition errors	Spatial span: backward total	Spatial span: backward length	Spatial span: combined	TMT: combined time	TMT: combined errors	TMT: contrast
CHEXI Working Memory	.137	.209	**.552****	**.500***	**.529***	.284	-.013	.095
CHEXI Inhibition	.162	.260	**.522***	**.527***	**.506***	.374	-.128	.113

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

3.7 Acceptability

Descriptive statistics for the visual analogue scale can be seen in Table 18. From this table, it can be seen that all six measures were broadly rated as acceptable. Wilcoxon's Signed-Rank test confirmed that there were no reliable differences between ratings of Dragon Adventure and ratings of the existing performance-based measures (all $p > .307$).

Table 19. Descriptive statistics for acceptability ratings on the novel and existing performance-based measures.

	Dragon Dash	Dragon Sequence	Dragon Hunt	CWIT	Spatial Spans	TMT
Mean	4.02	3.62	3.95	3.95	3.74	3.64
SD	0.72	0.87	1.21	0.90	1.24	0.89

4. DISCUSSION

This section begins by providing a summary of the research findings. The research aims and questions will be revisited, alongside discussion of the relevant results. This section also provides a discussion of present results in relation to the wider literature. The critical evaluation is then presented, followed by the clinical implications, directions for future research and final summary.

4.1 Summary of Findings

4.1.1 Revisiting the Aims and Research Questions

The aim of the present study was to establish the feasibility of assessing executive functioning in children using novel, game-like, computerised tasks. Specifically, the study aimed to develop a computerised novel measure that was embedded within a three-dimensional space and convincingly replicated a game.
The study aimed to address the following research questions:

- Is Dragon Adventure more engaging for children than standardised measures of executive functioning?
- Does performance on Dragon Adventure correlate with standardised measures of executive functioning (concurrent criterion validity)?
- Does performance on Dragon Adventure correlate with a teacher-rated measure of executive functioning (predictive criterion validity)?

4.1.2 Game-like procedure

As stated above, the study aimed to create a novel game-like computerised measure. To evaluate how successfully the current measure utilised a game-like format, Crawford’s (2012) criteria were used:

- **Rules** of play and underlying logic.
- **Interactivity** between the player and the game experience.
- **Immersion, engagement and flow**.
- **Performance** by experimenting with alternative identities.
- **Identify, roles and embodiment**.
- **Intertextuality and integration** with other media forms (e.g., film, novels)

- **Narrative**, telling or generating story.
- **Geography**, the space in which the game takes place.

In each of the three main tasks in Dragon Adventure, the participant was presented with a novel set of rules about how to complete the task. These can be considered operational rules, for example “avoid the obstacles to get to the end”. Additionally, the game included some constitutive rules, for example the background program that allows the character to jump.

Dragon Adventure included some aspects of interactivity, for example the participant is able to navigate around the environment relatively autonomously. However, the participant has limited ability to alter the course of gameplay, for example the main character will always ‘succeed’ and collect a reward gem, regardless of performance. This was somewhat necessitated by the requirement that it also function as a neuropsychological test. For example, it would not have been possible for the participant to ‘fail’ and then attempt the task again as can be done in video games. The researcher chose to reward each participant with a gem in order to prevent participants from becoming demotivated, although this may have increased a sense that the player does not have an impact on the outcome of the task. Despite this, participants did receive feedback if they made a mistake on an individual trial, for example an ‘incorrect’ sound if a participant tried to pick up the wrong item on the Dragon Hunt task. The ‘immersion, engagement and flow’ criteria are fully discussed in section 4.1.3.

Dragon Adventure provides the opportunity for the participant to ‘perform’ the identity of the main character (a dragon attempting to navigate home). The participants were able to experiment with ‘being’ a dragon character, for example by flying or breathing fire. This also enabled the participants to engage in an alternative ‘identity’ and play the role of that character. This could have been enhanced by enabling the player to have some level of interactivity with the main character, for example by choosing the colour, age or name of the dragon. The performative nature was limited to ‘performing to self’, whereas multiplayer video games also enable the player to ‘perform to others’.

Dragon Adventure does not explicitly make use of intertextuality, (i.e., it is not an extension of a story told through other media). It does, however, make reference to the fantasy genre and provides the participants with a sense of familiarity. For example, dragons have been used in many novels, films and games recently as well as being routed in longstanding cultural narratives (Sheridan, 2015).

There is an overarching narrative in Dragon Adventure, (i.e., a dragon character is lost and needs to collect ‘gems’ to navigate home). The Dragon Adventure game is periodically interrupted to update the participant on the status of the narrative (e.g., “you are nearly home”). The narrative element could be further developed by encountering third-party characters, thickening the story or using animation clips.

Finally, Dragon Adventure provides a relatively large environment (Appendix A –G) in which the ‘game’ takes place. The visual appeal of the geography is enhanced by the art assets available in the Unity Asset Store (Unity Technologies, 2020). The geography could be improved by increasing the interactivity with objects or by creating a ‘world map’ for the participant to reference.

In summary, Dragon Adventure appears to meet more of Crawford’s (2012) criteria than both Johann and Karbach (2018) and Józsa, Barrett and Morgan (2017). This suggests that Dragon Adventure is more ‘game-like’ than these previous measures. Nevertheless, Dragon Adventure is in an early stage of development and there are areas for improvement.

4.1.3 Research question 1

The study aimed to investigate whether Dragon Adventure was more engaging for children than standardised measures of executive functioning. In order to measure engagement and acceptability, the participants were asked to rate how enjoyable they found each task on a scale of 1 – 5 using a visual analogue scale (Appendix H). Wilcoxon’s Signed-Rank tests were used to compare the acceptability ratings on the novel and existing performance-based measures.

4.1.3.1 Interpretation of findings

The participants rated all the tasks as broadly acceptable and there was no reliable difference between the ratings of Dragon Adventure and those of the existing measures. The results indicate that the children found both Dragon Adventure and the existing performance-based measures to be enjoyable and engaging.

Whilst it may initially appear to undermine the concept that games are a potentially useful approach to increasing engagement neuropsychological testing, it is important to consider that Dragon Adventure was presented at an early stage of development. It therefore lacked some of the more advanced features that a typical video game contains (e.g., advanced scoring systems, high levels of intractability, customisable elements, developed storylines, animation scenes; Crawford, 2012). Engagement and acceptability may also have been limited by some of the points identified in section 4.1.2. For example, the game progressing regardless of performance on the tasks. The participants may have been comparing their enjoyment of Dragon Adventure with established video games, potentially rating the game lower by comparison.

Anecdotally, many participants volunteered suggestions for how to improve the game-like elements of Dragon Adventure (e.g., customisable main character, interactable objects, inventory systems). This highlights the advantages of involving children and young people in research (Greig, Taylor, MacKay, 2012), something that was not utilised in the present study due to time constraints. Consulting with or co-creating Dragon Adventure with young people would have provided the opportunity to integrate this information in the test development stage.

It is also important to consider possible areas of bias. Given the small range in mean acceptability ratings (3.62 – 4.02), it appears that ratings may have been influenced by acquiescence (Barker, Pistrang & Elliot, 2002). Recruitment took place within the secondary school that the participants attended. This, in combination with the status of the adult researcher, may have led to power differentials (Hagan & Smail, 1997) that prevented some participants from providing ‘honest’ ratings and most participants agreeing with the statement “I enjoyed the task”. Providing a way for participants to respond with anonymity may be a way to avoid this in the future.

4.1.4 Research Question 2

The second research question was concerned with concurrent criterion validity. The study aimed to investigate whether performance on Dragon Adventure correlated with performance on existing measures of executive functioning. The participant’s scores on Dragon Adventure were compared with scores on the D-KEFS Colour-Word Interference test (CWIT; Delis et al., 2001), WNV Spatial Span (Wechsler & Naglieri, 2006) and the D-KEFS Trail Making Test, (TMT; Delis et al., 2001). Spearman’s rank correlation coefficients were used to examine relationships between the measures.

4.1.4.1 Interpretation of findings

A number of moderate-to-large correlations were observed between the novel measures and the established performance-based measures.

On the Dragon Dash measure, the results indicated that as errors increased so did the time it takes participants to complete the CWIT. Errors on Dragon Dash were also associated with the TMT, participants who made more errors on Dragon Dash also took longer to complete the TMT – N-L-S task.

These results suggest that Dragon Dash may be successfully measuring some aspect of the construct of inhibition. The association with the TMT may be indicative of the link between inhibition and flexibility (i.e., that you must successfully stop one task to be able to switch to another) (Miyake et al., 2000). It is encouraging that Dragon Dash is moderately correlated with completion time on the CWIT however there was no association with errors on the CWIT. It is not clear why this is the case but may perhaps indicate that there are differences in verbal inhibition on the CWIT and inhibition in the visual-spatial domain on the Dragon Dash measure. Indeed, there is some evidence that children’s performance on Go/no-go tasks and colour-word Stroop tasks are only weakly correlated and follow different developmental courses, with performance on the Stroop improving at a younger age than the Go/no-go task (Morooka, 2012). Even within paradigms that are similar, such as the

Go/no-go and Stop-signal tasks, there is evidence to suggest different underlying processes (Littman & Takács, 2017). These results highlight the problem of assuming that inhibition is a unified construct (MacLeod, 2007) and the importance of understanding which cognitive processes are employed in tasks that are apparently similar.

The novel Dragon Sequence task was associated with several scores on the existing measures. It was most strongly associated with the Spatial Span – Backwards task. The total number of correctly recalled items on Dragon Sequence was strongly associated with the total correct and moderately associated with the longest correct scores on Spatial Span. Both the total correct and errors per attempt made on the novel Dragon Sequence task were associated with the combined Spatial Spans score. The Dragon Sequence task was also associated with the CWIT, with both the total correct and errors per attempt strongly associated with the CWIT completion time. Additionally, errors per attempt was moderately associated with CWIT errors. There were no associations with the TMT.

It is perhaps unsurprising that the novel Dragon Sequence task was most strongly associated with the Spatial Span measure. Of all the novel tasks it is the one that appears to most closely resemble the established measure. This suggests that Dragon Sequence may have been successful in measuring the visuo-spatial aspect of working memory. The strong association with the CWIT is less straightforward to interpret, especially considering that associations between the WNV Spatial Span measure and the CWIT, as well as the novel Dragon Sequence and Dragon Dash measures were all small. It seems that, perhaps in addition to working memory, the novel Dragon Sequence task may also measure an unknown skill common to both Dragon Sequence and the CWIT. For example, it may address the distribution of visual attention, scanning of a scene and sequencing a response.

The Dragon Hunt task had some associations with all three established tests. Both the completion time and errors on the Dragon Hunt task were strongly associated with the CWIT completion time, with good performance on one task indicating good performance on the other. Additionally, both completion time and errors on the Dragon Hunt task were moderately associated with Spatial Spans longest correct.

Finally, errors on the Dragon Hunt task was moderately associated with completion time on the TMT but all other associations were small.

The fact that the novel Dragon Hunt task had some association with all three established tests may indicate that this task was recruiting more aspects of EF than the other novel tasks. For example, to succeed on the Dragon Hunt task participants need to remember which object was picked up last (as there is no visual prompt as with the TMT). They also need to inhibit the previous response (e.g., stop picking up blue objects) and switch to a new response (e.g., now pick up a red object). The lack of visual prompts may explain why associations with the TMT were not stronger. It could be that the added demand on memory stores means the Dragon Hunt task is less able to isolate the flexibility skill. Future versions of the Dragon Hunt measure may therefore be improved by adding trails with and without visual prompts.

Finally, the different associations between each of the novel measures and the established measures indicate that the three tasks seem to address different abilities.

4.1.5 Research Question 3

The third research question was concerned with predictive criterion validity. The study aimed to investigate whether performance on Dragon Adventure correlated with a teacher-rated measure of executive functioning. To assess this, the scores on Dragon Adventure were compared with scores on the CHEXI (Thorell & Nyberg, 2008). Spearman's rank correlation coefficients were used to examine relationships between the measures.

4.1.5.1 Interpretation of findings

The CHEXI was most strongly associated with the novel Dragon Sequence task, with strong associations found between both the Working Memory and Inhibition subscales. This suggests that the Dragon Sequence task may be a useful predictor of teacher-rated executive functioning skills in children. The remaining associations were small, although the largest of these were between the Dragon Dash task and both subscales on the CHEXI. Camerota, Willoughby, Kun and Balir (2016) found only very small correlations (r = -.10 and -.05) between both subscales on the CHEXI and a computerised EF battery in a large sample of children aged 3-5 years. In a sample of 202 adults using the Adult Executive Functioning Inventory (ADEXI) Holst and Thorell (2018) found small but significant associations between the working memory subscale and WAIS-IV Digit span, WAIS-IV Number-letter sequencing and the D-KEFS CWIT test. Similar sized, and larger, correlations were found between Dragon Adventure and the CHEXI in the current study.

Thus, the results for Dragon Adventure are in line with, and perhaps slightly better than, previous studies comparing the Executive Functioning Inventory with performance-based measures of EF. Although, results for Dragon Adventure were similar to the associations found on the established measures in the current study. Overall, it suggests that self-report and teacher-reported measures of EF may be measuring different, although related, constructs (Toplak et al., 2013). Similar results have been found using the BRIEF, another widely used questionnaire-based measure of EF (Fernández et al., 2014; Silver, 2014; Toplak et al., 2013). Incorporating additional indications of EF performance in daily life (e.g., academic records) may be useful.

In summary, the three novel tests have positive outcomes for predictive criterion validity.

4.1.6 Additional Findings

In addition to the research questions, the differences between sex and language groups were examined along with associations within Dragon Adventure and internal consistency. These are detailed below.

4.1.6.1 *Sex and language groups*

Male participants scored slightly better on all variables apart from the Dragon Hunt task where they made more errors than female participants. The only statistically reliable difference was for the Dragon Sequence task, where males were able to remember longer sequences than female participants. This difference may be a chance finding or may reflect a greater level of familiarity with the video-game format in males compared to females (Griffiths, Davies & Chappell, 2004; Paulus, Ohmann, Gontard & Popow, 2012; Williams, Yee, Caplan, 2008). The difference on the Dragon Sequence task may be most apparent due to the duration of the task, which was comparatively shorter than the other measures. This may have meant the participants had less time to learn and understand the task, which for female participants may have been more pronounced due to having less prior knowledge of the video-game format. The likelihood of male participants having greater levels of familiarity with the format of Dragon Adventure is important to consider for future research, to ensure that female participants are not disadvantaged. Dragon Adventure may be improved by including practice items. These were not included in the current measure due to time constraints but would allow participants the opportunity to learn the rules of the task without being penalised for initial mistakes. Establishing separate norms for males and females may also be beneficial.

The EPL group also scored slightly better on all variables apart from the Dragon Hunt task, where they made more errors than the EAL group. There were two statistically significant differences. Firstly, on the Dragon Dash task where the EPL group made fewer errors than the EAL group. Secondly, on the Dragon Sequence task, where the EPL group were able to remember longer sequences than the EAL group. The test instructions were presented in written English, which may explain some of these differences. Although all participants were able to read and speak English, it may still have created disadvantage to these participants. Previous studies have found mixed results as to whether performance on neuropsychological tasks is influenced by primary or additional language status, but performance on verbal tasks may be negatively impacted (Kisser, Wendell, Spencer & Waldsetin, 2012). Dragon Adventure may be improved by presenting the instructions in a non-verbal format, for example by video clips that demonstrate play. As stated above, practice trials may also be beneficial.

4.1.6.2 Motor speed

No association was found between the TMT number sequencing time scores and the novel measures. This suggests that motor speed may have played a limited role participants' performance on Dragon Adventure. However, the TMT is not primarily used as a measure of pure motor speed and the pen and paper task may not be applicable to motor speed on a computer-based task. Motor speed is likely to play a role in all three of the Dragon Adventure tasks and large variations in completion time can be seen on Dragon Hunt in particular (3.03 minutes – 17.05 minutes). At present it is not possible to distinguish general motor speed from EF skills and future research should include a motor speed measure within Dragon Adventure.

4.1.6.3 Associations within Dragon Adventure

Within each of the novel tests, the largest association was between two variables on the Dragon Sequence task. This was the total number of sequences correctly answered and the longest sequence correctly recalled ($r = .978$). This suggests that, within this sample, *Longest Correct* and *Total Correct* could be considered to be measuring the same ability. The number correctly recalled on Dragon Sequence was also strongly negatively associated with the errors per attempt. On the Dragon Hunt task, there was a strong positive correlation between the completion time and error rate variables. Overall, the strong associations within both the Dragon Sequence and Dragon Hunt tasks indicate that the different variables are measuring similar constructs, purportedly working memory and flexibility respectively.

When comparing between the novel tasks, Dragon Dash was strongly associated with both errors and completion time on the Dragon Hunt task. Dragon Dash was also moderately associated with Dragon Sequence – *Longest Correct*, and there was a small association with *Total Correct*. This indicates that the Inhibition task is measuring a skill that is related to those that are used to succeed at the Dragon Sequence and Dragon Hunt task. This is in line with modelling of EF, which suggests that these constructs are overlapping (Miyake et al., 2000). There was no, or very small, associations between the Dragon Sequence and Dragon Hunt tasks. The lack of association between these tasks suggests that the tasks are measuring relatively independent skills, and importantly, that performance on all tasks are not primarily influenced by an underlying universal skill (e.g., computer literacy or video-game familiarity).

4.1.6.4 Internal consistency

As part of the process of test development, Cronbach's alpha and Spearman-Brown Coefficients were used to examine the internal consistency of Dragon Adventure. The Dragon Sequence task was found to have the greatest reliability, followed by the Dragon Hunt task and finally the Dragon Dash task. The Dragon Sequence task and the Dragon Hunt task were found to have good reliability and Dragon Dash was found to have acceptable reliability. Overall, this indicates that, within each task, the same underlying construct is being measured.

For the Dragon Sequence task, deleting the first item resulted in a small increase in reliability. Although small, this finding supports the suggestion to introduce practice trials on the Dragon Sequence task. It is possible that the first trial is also measuring another underlying concept, such as familiarity with the rules. Practice trials may provide the opportunity for participants to learn the rules of the task, thus potentially increasing the reliability for the remainder of the task.

4.1.7 Summary

The present study aimed to create a convincingly game-like measure of EF for children. Dragon Adventure met a number of Crawford's (2012) criteria for video games and more so than previous literature. Nevertheless, the measure is in an early stage of development and several areas for development were identified.

The first research question asked if Dragon Adventure was more engaging than existing performance-based measures of EF. Dragon Adventure was found to be no more engaging that the existing measures and all were rated as broadly 'enjoyable'. This may be due to the early stage of development of Dragon Adventure, but acquiescence may also play a role.

The second research question asked if Dragon Adventure correlated with existing performance-based measures of EF (concurrent criterion validity). A number of moderate-to-large correlations were found between the novel and established measures, indicating that Dragon Adventure may be successfully measuring EF.

The third research question asked if Dragon Adventure correlated with a teacher-rated measure of EF (predictive criterion validity). There was a strong association between the Dragon Sequence task and both subscales on the CHEXI. All other associations were small. Associations with performance-based and report-based measures of EF are often poorly associated, although the results from the current study appear to be slightly better connected than previous literature findings.

Males scored slightly better on most variables on Dragon Adventure, although only one difference was significant (*Dragon Sequence – Total Correct*). Similarly, the EPL group performed slightly better on most variables, with two significant differences (*Dragon Dash – Total Errors* and *Dragon Sequence – Longest Correct*). These results suggest that Dragon Adventure needs to be revised to reduce potential disadvantage to these groups, for example by including practice trials and non-verbal demonstrations of the task. Establishing separate norms for males and females may also be beneficial. Dragon Adventure cannot distinguish between general motor speed and EF skills at present, future research should incorporate a motor speed test in Dragon Adventure.

There were a number of associations both between and within Dragon Adventure. Results indicate that within the subtest, variables are measuring similar constructs, and between variables there is some expected overlap given the interrelated nature of EF. Finally, all measures were found to have good to acceptable internal consistency.

4.2 Critical Evaluation

4.2.1 Strengths

The current study is one of the first to use a three-dimensional game-like computerised format to measure EF in children. To the researcher’s knowledge, the Johann and Karbach (2018) study, Józsa, Barrett & Morgan (2017) study and the current project, are the only studies to begin using this format to assess EF in children. There are numerous other studies that have investigated the use of game-like paradigms in neuropsychological testing (see Lumsden, Edwards, Lawrence, Coyle & Munafò, 2016, for a review), however they have been concerned with different age ranges, specialist populations (e.g., ADHD, Alzheimer's disease) or other cognitive functions. This study has added to the literature by enhancing the number of game-like components and utilising a three-dimensional game environment.

Similarly, it is one of the first studies to use the Unity software in the field of psychology. Recently, it has been identified as a useful tool for virtual reality experimental tasks (e.g., Brookes, Warburton, Alghadier, Mon-Williams & Mushtaq, 2020; Rizzo, Hartholt, Grimani, Leeds & Liewer, 2014; Wiesing, Fink & Weidner, 2020) and although not used in the current study, Brookes et al. (2020) have created The Unity Experiment Framework to bridge the gap between sophisticated game engines and behavioural scientists. The Unity software offers the opportunity to significantly enhance both game-like features of assessment and ecological validity through the art assets. It provides an opportunity for cross-discipline research by collaboration with an extensive network of established game developers and students.

Using the Unity software enabled the researcher to create a 'novel' environment (i.e., the game-space) for the testing to take place and therefore provides an opportunity to create a more 'culturally fair' measure. Whilst there are issues with participants' level of familiarity with video-game paradigms and language use for test instructions, it significantly reduces the possibility of cognitive testing mimicking an educational environment and therefore benefitting those with more experience of education systems (Fernandez & Abe, 2018). Using a computer format also allows the researcher to collect a large amount of data and minimise errors that can occur during pen-and-paper administration (Goldstein & Naglieri, 2013).

Finally, the literature review identified that many tools used to assess EF in children were originally designed for use in adult populations (Anderson, P., 2002; Anderson, V., 2001; Anderson, V., 2002; Archibald & Kerns, 1999). A strength of the current study is that it was designed specifically for use in a child population. The results discussed above indicate that, despite the measure being in an early stage of development, it has the potential to be an effective and reliable measure of EF in children.

4.2.2 Limitations

Despite the strengths discussed above, it is important to consider the limitations of the current study. Firstly, the sample size was relatively small (N = 21) meaning it was not possible to make conclusions about the normative performance on Dragon Adventure. Additionally, the study may have lacked statistical power to examine differences between the groups, for example between the performance of male and female participants. The EAL group was particularly small (N = 5; 23.8%), this is compared to 16.9% nationally and 55.6% in inner London (Department for Education, 2019). The percentage of pupils with EAL is therefore significantly lower than the inner London average. This may in part be because the ability to read and write English (sufficient to read the instructions) was an inclusion criterion. Additionally, as consent and information forms were written in English, it is possible that some parents and guardians were unable to read the information resulting in lower return rates. Caution should be taken when interpreting differences in performance for this group, both because of the very small sample size and the under-representation of EAL pupils in London. Future research should use a larger sample to minimise these issues. As the main aim of the current study was to create Dragon Adventure and begin exploring the feasibility of measuring EF with this tool, it was anticipated at the start of the project that larger-scale studies would be an appropriate next step if indicated by this study.

Interpretations of the results are limited by not knowing how familiar the participants were with video games. To gain a clearer understanding of whether, or how much, familiarity influences performance, participants should be asked how often they play video games, what type (e.g., action, adventure, simulation) and using which format (e.g., mobile phone, computers, consoles). It would then be possible, if the sample size is sufficient, to model the contribution of familiarity with video games to test performance.

Similarly, the level of familiarity with computers and/or laptops was not measured. During testing it became apparent that two participants did not know how to use a keyboard and needed additional instructions by the researcher on how to use the keys. Although the initial instruction scene provided an opportunity to practice keyboard skills, during the Dragon Hunt task these two participants took far longer to complete the task than their peers. This highlights the role of computer knowledge and fine motor skills in performance on Dragon Adventure. Whilst it is not possible to create a measure that does not rely on any lower-order skills, it remains important to consider the influence of such skills.

The present study also did not include a measure of qualitative factors in cognitive assessment (V. Anderson, 2002). In clinical practice, it is important to consider the influence of a range of non-cognitive factors (e.g., motivation, attention, strategy, emotional state) on performance outcome. Although a measure of engagement was included in the study, additional measures (e.g., self-report measures of emotional state) could have been included to assess the potential contribution of these additional factors. As with all neuropsychological testing, it remains important that the clinicians interpreting the test results are suitably trained in order to integrate contextual and cognitive information into a comprehensive formulation.

In terms of the design of Dragon Adventure, several areas for improvement have been identified above. Perhaps the most important are those concerning the children's understanding of test instructions. None of the tasks included practice trials and all relied on English to communicate the instructions. In addition, some participants needed the instructions repeated verbally by the researcher as they had clicked through all the instructions before reading thoroughly. The instructions could therefore be improved by allowed the participants to press a 'back' button, to re-read instructions. As discussed above, including practice trials and demonstrating instructions via video clips would go some way to ensuring that all participants understood the instructions and minimising language bias.

Regarding engagement, the study was limited by the visual analogue tool being administered with the researcher present. This may have prevented some participants from disagreeing with the statement "I enjoyed the task". Indeed, anonymity is often cited as an important aspect for those who are giving feedback (Speed, Davison & Gunnell, 2016). In addition, many participants asked questions about Dragon Adventure and who made it. These participants therefore knew that the researcher had created Dragon Adventure, and this may have prevented some from rating it low for enjoyment. It is currently unclear how much of an influence this may have had, future research would benefit from providing anonymised feedback methods.

4.2.3 Personal Reflections

This study adopted a critical realist position (Danermark, Ekström, Jakobsen & Karlsson, 2002; Willig, 2012), which understands that research data is influenced by the context in which it was generated. As such, research reflexivity (Flanagan, 1981) is an important component of the work. Being reflexive in research requires the researcher to discuss the context in which the research arose and the relationship between this, the researcher and the research results.

The decision to adopt a critical realist approach was one of the first choices made in the process. I feel that this represents an integration of two large influences in my personal life. Firstly, my family whom tend to align with a positivist position and value reason and logic as methods of gaining knowledge. Secondly, the doctoral training which has provided extensive thought and critical reflection on the issues of clinical psychology and epistemology. At times during the process I have noticed that I had slipped into more of a positivist position. For example, by considering the construct of EF as objectively measurable, and questioning whether Dragon Adventure is capable of directly observing EF domains. During the research project it has therefore been helpful to acknowledge that the construct of EF is socially constructed, thus making measurement an attempt to observe behavioural output that exists in context.

Throughout the research project I have wondered whether the research may unintentionally contribute to a medicalised and de-contextualised understanding of cognitive function. Describing performance on EF test as ‘better’ or ‘worse’ lends itself to a deficit model of understanding differences in EF. This model risks these differences as being viewed as held within the person rather than contextual. For example, ADHD has become a widely used diagnostic label associated with deficits in EF in children and has been widely criticised for being Westernised and decontextualized (Timimi & Timimi, 2015). Nevertheless, I feel that striving to better understand cognitive function, and the contextual factors that influence development and expression of these functions, will ultimately be a useful endeavour. Having reliable and valid tests for children is important, not only so that 'deficits' are not inappropriately diagnosed, but so that children who are experiencing difficulties in their daily life can be given appropriate support.

During the testing and data collection I was struck by the position I was placed in as an adult in the school. The students were expected to call me "madam" and open doors for me as we walked to and from the testing room. This is a stark contrast to the way in which I try to practice in a clinical environment, taking a one-down position in an attempt to lessen the effects of power. I wondered how being in a position of relative authority influenced the results of the study, particularly on the ratings of 'enjoyment', as most participants rated all elements favourably. I also feel that experiencing this different positioning has been useful in highlighting the issue of power my research and clinical practice. Taking a one-down position can lead to a false sense of equal power sharing and therefore inadequate reflection on the role of power by the clinician. Experiencing this obvious power differential has reminded me that the effects of power cannot be eradicated but instead must be thoughtfully considered, named and counter-balanced where possible.

Lastly, developing Dragon Adventure has been a difficult task and took much longer to complete than I anticipated at the beginning of the process. It involved learning a substantial amount of new information, specifically the coding language C# and the Unity engine software. Undertaking this new learning, in addition to the learning in the professional doctorate training has been greatly challenging at times. I wonder whether my decision to take on the task was linked with my stage of training and discomfort in grappling with concepts that are not clear cut. Learning how to create a computer game, with clear 'right' or 'wrong' answers, provided some comfort that perhaps acted as a counterbalance for the less tangible learning on the doctorate. It also acted as a way to 'escape' some aspects of the academic and therapeutic work into a discipline that was new to me, whilst still working towards qualification. Whilst creating Dragon Adventure has been challenging, it has also been decidedly satisfying. Despite this, creating Dragon Adventure whilst contending with multiple other academic and placement demands is likely to have led to compromises in the design of the study. Perhaps the largest compromise was not involving young people in the creation of Dragon Adventure. This was not done, in favour of a process whereby decisions could be made quickly within a task-oriented framework.

As the process of completing the thesis draws to a close, the world is coming to terms with life during the COVID-19 pandemic. I feel that, in a similar way that this work provided comfort during training, it has now also provided comfort and consistency during a very uncertain time both for the NHS and globally.

4.3 Clinical Implications

The findings of the study suggest that, with development, Dragon Adventure could be a useful tool to measure EF children. Developing Dragon Adventure into a neuropsychological test appears to be a feasible project and designing an assessment tool specifically for children also provides opportunities to improve engagement and validity. The study also contributes to a growing body of research that is focused on developing appropriate neuropsychological test for children. Developing a reliable and valid test of EF in children will improve understanding and intervention for children who are experiencing difficulties in their daily life linked to EF skills.

An effective test would also improve understanding of the developmental trajectory of EF in children. This has only recently been viewed as an issue of concern (Anderson, P., 2002) and current research is limited by the challenges of assessing EF in children (see section 1.5.2). Improving tests of EF for children will allow researchers to develop a model of typical development, and therefore more easily identify children whose EF development has deviated from the typical pattern. This would also support accurate identification of difficulties versus ‘normal’ variants in development. A developmental model would also complement research in EF difficulties in other populations (e.g., adults, brain injury).

As the measure uses a game-like format and was rated as ‘enjoyable’, there are opportunities for clinicians to build engagement with children. This may be particularly beneficial for children who have suspected difficulties with EF. This group is more likely to have had difficulties in a school environment compared to peers (Biederman, et al., 2004) and may therefore find traditional testing measures anxiety-provoking or unappealing due to the resemblance to a school task or test. The game-like format can provide an engaging alternative for clinicians to build rapport. If children are more relaxed during the testing, clinicians are also more likely to gain an accurate representation of their skills (Chaytor & Schmitter-Edgecombe, 2003).

Finally, as the entire measure is administered on the computer, Dragon Adventure has the potential to be used widely. In addition, scoring Dragon Adventure could be developed to be completed automatically thus reducing scoring errors. These features mean that the novel could be used by a variety of settings (e.g., schools, clinics, home-based) and could save time in busy psychological services. Nevertheless, interpretation of the results would need to be completed by an appropriately trained clinician.

4.4 Recommendations and Future Research

This study was the first stage in the development of a new game-like measure of EF for children. The results indicate that Dragon Adventure has the potential to be a useful tool in measuring EF in 11 to 12-year-olds. Future research should focus on further development of the measure, assessing reliability and validity, and developing normative data. The recommendations for test design are collated below in section 4.4.1.

To assess reliability and begin gathering normative data, Dragon Adventure should be administered to a larger sample. A larger sample size would allow the researcher to assess whether there are differences in performance between groups. For example, differences between male and female groups and EPL and EAL groups. Future studies should seek to include a larger sample of children with EAL and ensure that inclusion criteria and sampling methods capture a representative sample. Dragon Adventure should also be administered to a wider age range of children to assess the suitability of the test for younger and older ages. Future studies should also assess test-retest reliability by administering Dragon Adventure with the same group of children over two time points. This would allow the researchers to investigate the level of consistency between the two time points and the role of practice effects.

To measure engagement and assess the role of bias in the ‘enjoyment’ ratings of the novel and existing measures, future studies should use an anonymised (or semi-anonymised) collection method. For Dragon Adventure, this could be done by integrating the rating into the measure itself. Participants could then rate the measure on screen, without direct involvement from the researcher. Alternatively, participants could rate the measure with an alternative person, someone who is not the main researcher.

Future research should also include a measure of both video game familiarity and level of computer literacy. It would then be possible to assess the role of these factors in overall task performance.

To assess criterion predictive validity, additional indicators of EF skills should be included. The current study was limited to teacher-ratings of EF, therefore mainly reflective of the classroom environment. Future studies would benefit from gathering data about a child’s EF skills in daily life (e.g., parental ratings of functioning outside of the school environment) as EF has significant implications for a child’s social and emotional functioning as well as school academic performance (Best, Miller & Jones, 2009). To measure these aspects of executive functioning previous research has used parent-reports, academic reports as well as measures of emotional and social functioning (see Best, Miller & Jones, 2009 for a review). Therefore, future studies could use the CHEXI (Thorrell & Nyber, 2008) alongside other measures of daily functioning, for example the Behaviour Rating Inventory of Executive Function (BREIF; Gioia et al., 2000), the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001) or school behavioural and academic records.

Lastly, future research could recruit participants who have ‘known’ difficulties with EF (e.g., ADHD, ASD, brain injury). It would then possible to assess whether Dragon

Adventure is capable of discriminating between groups of children with a ‘known’ EF difficulty and those who do not. It would also provide an opportunity to assess whether Dragon Adventure is reliable and valid in a population of children with EF difficulties.

4.4.1 Test Design

Several areas for improvement for the design of Dragon Adventure were identified in this discussion. For clarity, these are collated here:

- Improve the game-like element of Dragon Adventure:
    - Consulting with children regarding the design of Dragon Adventure.
    - Improve interactivity with the main character, for example by allowing the child to change the colour, age or name of the dragon.
    - Develop the narrative element of Dragon Adventure by thickening the story, utilising third-part characters and/or using animation clips.
    - Increasing interactivity with objects within the world.
    - Creating a ‘world map’ for game environment.
- Create an anonymous feed-back process for children to rate their ‘enjoyment’ of the task.
- Include practice trials for all tasks on Dragon Adventure.
- Include a test of motor speed.
- Include video-clips that demonstrate test instructions for all tasks.
- Include a ‘back’ button for test instructions, so children can re-read if needed.
- Consider including trails with and without visual prompts on the Dragon Hunt task to vary demand on working memory.

5. Concluding Summary

This study investigated the feasibility of using a novel game-like measure to test executive functioning in children. The results indicate that Dragon Adventure has fulfilled a number of criteria for a video game and was rated as ‘enjoyable’ by participants. Dragon Adventure was correlated with established measures of executive functioning and showed the potential to be an effective and reliable tool for measuring executive functioning in children. Future research can now be conducted to improve the design of Dragon Adventure, assess engagement, and develop it into a reliable and valid neuropsychological measure.

REFERENCES

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, *8*(2), 71–82. https://doi.org/10.1076/chin.8.2.71.8724

Anderson, P., Anderson, V., Northam, E., & Taylor, H. G. (2000). Standardization of the Contingency Naming Test (CNT) for school-age children: A measure of reactive flexibility. *Clinical Neuropsychological Assessment*, *1*, 247–273.

Anderson, V. (2001). Assessing executive functions in children: Biological, psychological, and developmental considerations. *Pediatric Rehabilitation*, *4*(3), 119–136. https://doi.org/10.1080/13638490110091347

Anderson, V. (2002). Executive function in children: Introduction. *Child Neuropsychology*, *8*(2), 69–70. https://doi.org/10.1076/chin.8.2.69.8725

Archibald, S. J., & Kerns, K. A. (1999). Identification and description of new tests of executive functioning in children. *Child Neuropsychology*, *5*(2), 115–129. https://doi.org/10.1076/chin.5.2.115.3167

Aron, A. R., Robbins, T. W., & Poldrack, R. A. (2014). Inhibition and the right inferior frontal cortex: One decade on. *Trends in Cognitive Sciences*, *18*(4), 177–185. https://doi.org/10.1016/j.tics.2013.12.003

Baddeley, A. & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8*, 47-89. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839. https://doi.org/10.1038/nrn1201

Bari, A., & Robbins, T. W. (2013). Inhibition and impulsivity: Behavioral and neural basis of response control. *Progress in Neurobiology*, *108*, 44-79. https://doi.org/10.1016/j.pneurobio.2013.06.005

Barker, C., Pistrang, N., & Elliot, E. (2002). *Research methods in clinical psychology: An introduction for students and practitioners* (2nd ed.). West Sussex: John Wiley & Sons.

Bauer, P. J., & Zelazo, P. D. (2013). IX. NIH Toolbox Cognition Battery (CB): Summary, conclusions, and implications for cognitive development. *Monographs of the Society for Research in Child Development*, *78*(4), 133–146. https://doi.org/10.1111/mono.12039

Berg, E.A. (1948). A simple objective technique for measuring flexibility in thinking. *Journal of General Psychology*, 39,15-22. https://doi.org/10.1080/00221309.1948.9918159

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental review, 29*(3), 180-200. https://doi.org/10.1016/j.dr.2009.05.002

Biederman, J., Monuteaux, M. C., Doyle, A. E., Seidman, L. J., Wilens, T. E., Ferrero, F., ... & Faraone, S. V. (2004). Impact of executive function deficits and attention-deficit/hyperactivity disorder (ADHD) on academic outcomes in children. *Journal of Consulting and Clinical Psychology*, *72*(5), 757-766. https://doi.org/10.1037/0022-006X.72.5.757

Bissonette, G. B., Powell, E. M., & Roesch, M. R. (2013). Neural structures underlying set-shifting: Roles of medial prefrontal cortex and anterior cingulate cortex. *Behavioural Brain Research*, *250*, 91–101. https://doi.org/10.1016/j.bbr.2013.04.037

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., & Mushtaq, F. (2020). Studying human behavior with virtual reality: The Unity Experiment Framework. *Behavior Research Methods*, *52*, 455-463. https://doi.org/10.3758/s13428-019-01242-0

Bujoreanu, S. I., & Willis, G. W. (2008). Developmental and neuropsychological perspectives on the Wisconsin Card Sorting Test in children. *Developmental Neuropsychology*, *33*(5), 584–600. https://doi.org/10.1080/87565640802254364

Cambridge Cognition Ltd (2006). *Cambridge Neuropsychological Test Automated Battery (CANTAB).* Cambridge: Cambridge Cognition Ltd.

Camerota, M., Willoughby, M. T., Kuhn, L. J., & Blair, C. B. (2018). The Childhood Executive Functioning Inventory (CHEXI): Factor structure, measurement invariance, and correlates in US preschoolers. *Child Neuropsychology, 24*(3), 322-337. https://doi.org/10.1080/09297049.2016.1247795

Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. *Developmental Neuropsychology, 28*(2), 595-616. https://doi.org/10.1207/s15326942dn2802_3

Casey, B. J., Giedd, J. N., & Thomas, K. M. (2000). Structural and functional brain development and its relation to cognitive development. *Biological Psychology*, *54*(1-3), 241-257. https://doi.org/10.1016/S0301-0511(00)00058-2

Catale, C., Meulemans, T., & Thorell, L. B. (2015). The Childhood Executive Function Inventory (CHEXI): Confirmatory factorial analyses and cross-cultural clinical validity in a sample of 8–11 years old children. *Journal of Attention Disorders, 19*(6), 489-495. https://doi.org/10.1177/1087054712470971

Chan, A. Y. C., & Morgan, S-J. (2018). Assessing children's cognitive flexibility with the Shape Trail Test. *PloS One*, *13*(5), e0198254. https://doi.org/10.1371/journal.pone.0198254

Chan, R. C. K., Shum, D., Toulopoulou, T., & Chen, E. Y. H. (2008). Assessment of executive functions: Review of instruments and identification of critical issues.

Archives of Clinical Neuropsychology, *23*(2), 201–216. https://doi.org/10.1016/J.ACN.2007.08.010

Chaytor, N., & Schmitter-Edgecombe, M. (2003). The ecological validity of neuropsychological tests: A review of the literature on everyday cognitive skills. *Neuropsychology review*, *13*(4), 181-197. https://doi.org/10.1023/B:NERV.0000009483.91468.fb

Chelune, G. J., & Baer, R. A. (1986). Developmental norms for the Wisconsin Card Sorting Test. *Journal of Clinical and Experimental Neuropsychology*, *8*(3), 219–228. https://doi.org/10.1080/01688638608401314

Cohen, D. (2006). *The development of play*. UK, Hove: Routledge.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins. (2020). *Definition of 'game'.* https://www.collinsdictionary.com/dictionary/english/game

Cragg, L., & Nation, K. (2007). Self-ordered pointing as a test of working memory in typically developing children. *Memory, 15*(5), 526-535. https://doi.org/10.1080/09658210701390750

Crawford, G. (2012). *Video gamers*. New York, NY: Routledge.

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, *38*(9), 571–578. https://doi.org/10.1016/j.tins.2015.07.003

Danermark, B., Ekström, M., Jakobsen, L., & Karlsson, J. (2002). *Explaining society: Critical realism in the social sciences.* London: Routledge.

Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). *Delis-Kaplan Executive Function System, Technical Manual (D-KEFS)*. San Antonio: The Psychological Corporation.

Department for Education. (2019). *Schools, pupils and their characteristics: January 2018.* https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2018

Diamond, A. (2006). The early development of executive functions. In E. Bialystok & F. I. M. Craik (Eds.), *Lifespan cognition: Mechanisms of change* (pp. 70–95). Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195169539.003.0006

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Emslie, H., Wilson, F., Burden, V., Nimmo-Smith, I., & Wilson, B. A. (2003). *Behavioural Assessment of the Dysexecutive Syndrome for Children (BADS-C)*. London: The Psychological Corporation

Eriksen, B., A., & Eriksen, C., W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysic, 16*, 143-149. https://doi.org/10.3758/BF03203267

Fahy, J. K. (2014). Assessment of executive functions in school-aged children: Challenges and solutions for the SLP. *Perspectives on School-Based Issues*, *15*(4), 151–163. https://doi.org/10.1044/sbi15.4.151

Fernández, A. L., & Abe, J. (2018). Bias in cross-cultural neuropsychological testing: Problems and possible solutions. *Culture and Brain*, *6*(1), 1-35. https://doi.org/10.1007/s40167-017-0050-2

Fernández, T. G., González-Castro, P., Areces, D., Cueli, M., & Pérez, C. R. (2014). Executive functions in children and adolescents: The types of assessment measures used and implications for their validity in clinical and educational contexts. *Papeles Del Psicólogo*, *35*(3), 215–223.

Ferrier, J. F. (1854). *Institutes of metaphysics: The theory of knowing and being*. Edinburgh: Blackwood.

Fisher, C. J. (2006). *Validity of the Delis-Kaplan Executive Function system in pediatric populations* (Doctoral dissertation, George Fox University). Retrieved from: https://core.ac.uk/download/pdf/51118682.pdf

Fisk, J. E., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting and access. *Journal of Clinical and Experimental Neuropsychology, 26*(7), 874-890. https://doi.org/10.1080/13803390490510680

Flanagan, O. J. (1981). Psychology, progress, and the problem of reflexivity: A study in the epistemological foundations of psychology. *Journal of the History of the Behavioural Sciences, 17*(3), 375-386. https://doi.org/10.1002/1520-6696(198107)17:3<375::AID-JHBS2300170308>3.0.CO;2-U

Freeman, D., Haselton, P., Freeman, J., Spanlang, B., Kishore, S., Albery, E., ... & Nickless, A. (2018). Automated psychological therapy using immersive virtual reality for treatment of fear of heights: A single-blind, parallel-group, randomised controlled trial. *The Lancet Psychiatry*, *5*(8), 625-632. https://doi.org/10.1016/S2215-0366(18)30226-8

Friedman, N. P., & Miyake, A. (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex, 86,* 186-204. https://doi.org/10.1016/j.cortex.2016.04.023

Fuster, D., Cole, P., & Tan, P. (2008). The relation between executive functioning and emotion regulation in young children. *The Penn State McNair Journal*, *15*, 35–53.

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), 31–60. https://doi.org/10.1037/0033-2909.134.1.31

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, *40*(2), 177–190. https://doi.org/10.1037/0012-1649.40.2.177

Gerstadt, C. L., Hong, Y. J., & Diamond, A. (1994). The relationship between cognition and action: Performance of children 3 1/2 - 7 years old on a Stroop-like day-night test. *Cognition*, *53*(2), 129–153. https://doi.org/10.1016/0010-0277(94)90068-X

Gioia, G., Isquith, P., Guy, S., & Kenworthy, L. (2000). *BRIEF – Behaviour Rating Inventory of Executive Function, professional manual.* Odessa, FL: Psychological Assessment Resources.

Goldstein, S. & Naglieri, J. A. (2014). *Handbook of executive functioning.* New York: Springer.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1337-1345. https://doi.org/10.1097/00004583-200111000-00015

Greig, A., Taylor, J., & MacKay, T. (2012). *Doing research with children: A practical guide* (3rd ed.). London: Sage Publications.

Griffiths, M. D., Davies, M. N., & Chappell, D. (2004). Demographic factors and playing variables in online computer gaming. *CyberPsychology & Behavior*, *7*(4), 479-487. https://doi.org/10.1089/cpb.2004.7.479

Guare, R. (2014). Context in the development of executive functions in children. *Applied Neuropsychology: Child*, *3*(3), 226–232. https://doi.org/10.1080/21622965.2013.870015

Hagan, T., & Smail, D. (1997). Power-mapping—I. Background and basic methodology. *Journal of Community & Applied Social Psychology*, *7*(4), 257-267. https://doi.org/10.1002/(SICI)1099-1298(199709)7:4<257::AID-CASP428>3.0.CO;2-P

Harvey, L., Rose, M., Jonsson, R., & Lask, B. (2016). Age-related performance on the Brixton Spatial Anticipation test in healthy child and adolescent girls: A preliminary evaluation. *Applied Neuropsychology: Child*, *5*(1), 1–8. https://doi.org/10.1080/21622965.2013.814087

Heaton, R. K. (1981). *Wisconsin card sorting test manual.* Odessa, FL: Psychological Assessment Resources.

Henry, L. A., & Bettenay, C. (2010). The assessment of executive functioning in children. *Child and Adolescent Mental Health*, *15*(2), 110–119. https://doi.org/10.1111/j.1475-3588.2010.00557.x

Hofmann, W., Schmeichel, B. J., & Baddeley, A. D. (2012). Executive functions and self-regulation. *Trends in Cognitive Sciences*, *16*(3), 174–180. https://doi.org/10.1016/j.tics.2012.01.006

Holst, Y., & Thorell, L. B. (2018). Adult executive functioning inventory (ADEXI): Validity, reliability, and relations to ADHD. *International Journal of Methods in Psychiatric Research*, *27*(1), e1567. https://doi.org/10.1002/mpr.1567

Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology*, *28*(2), 617–644. https://doi.org/10.1207/s15326942dn2802_4

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*(2), 233-253. https://doi.org/10.1111/j.2044-835X.1998.tb00921.x

Huizinga, M., & van der Molen, M. W. (2007). Age-group differences in set-switching and set-maintenance on the Wisconsin Card Sorting Task. *Developmental Neuropsychology*, *31*(2), 193–215. https://doi.org/10.1080/87565640701190817

Huizinga, M., Dolan, C. V., & van der Molen, M. W. (2006). Age-related change in executive function: Developmental trends and a latent variable analysis. Neuropsychologia, 44(11), 2017-2036. https://doi.org/10.1016/j.neuropsychologia.2006.01.010

Johann, V. E., & Karbach, J. (2018). Validation of new online game-based executive function tasks for children. *Journal of Experimental Child Psychology*, *176*, 150–161. https://doi.org/10.1016/j.jecp.2018.07.009

Józsa, K., Barrett, K. C., & Morgan, G. A. (2017). Game-like tablet assessment of approaches to learning: Assessing mastery motivation and executive functions. *Electronic Journal of Research in Education Psychology*, *15*(3), 665–695. https://doi.org/10.14204/ejrep.43.17026

Jurado, M. B., & Rosselli, M. (2007). The elusive nature of executive functions: A review of our current understanding. *Neuropsychology Review*, *17*(3), 213–233. https://doi.org/10.1007/s11065-007-9040-z

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and reanalysis of latent variable studies. *Psychological Bulletin, 144*(11), 1147.http://dx.doi.org/10.1037/bul0000160

Kisser, J. E., Wendell, C. R., Spencer, R. J., & Waldstein, S. R. (2012). Neuropsychological performance of native versus non-native English speakers. *Archives of Clinical Neuropsychology*, *27*(7), 749-755. https://doi.org/10.1093/arclin/acs082

Klenberg, L., Korkman, M., & Lahti-Nuuttila, P. (2001). Differential development of attention and executive functions in 3-to 12-year-old Finnish children. *Developmental Neuropsychology*, *20*(1), 407-428. https://doi.org/10.1207/S15326942DN2001_6

Klimkeit, E. I., Mattingley, J. B., Sheppard, D. M., Farrow, M., & Bradshaw, J. L. (2004). Examining the development of attention and executive functions in children with a novel paradigm. *Child Neuropsychology*, *10*(3), 201–211. https://doi.org/10.1080/09297040409609811

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY II: A developmental neuropsychological assessment manual.* San Antonio: Psychological Corporation.

Lagattuta, K. H., Sayfan, L., & Monsour, M. (2011). A new measure for assessing executive function across a wide age range: Children and adults find happy-sad more difficult than day-night. *Developmental Science*, *14*(3), 481–489. https://doi.org/10.1111/j.1467-7687.2010.00994.x

Latzman, R. D., & Markon, K. E. (2010). The factor structure and age-related factorial invariance of the Delis-Kaplan Executive Function System (D-KEFS). *Assessment*, *17*(2), 172–184. https://doi.org/10.1177/1073191109356254

Lehto, J. E., Juujärvi, P., Kooistra, L., & Pulkkinen, L. (2003). Dimensions of executive functioning: Evidence from children. *British Journal of Developmental Psychology*, *21*(1), 59-80. https://doi.org/10.1348/026151003321164627

Lezak, M. D., Howieson, D. B., Bigler, E. D., & Tranel, D. (2012). *Neuropsychological assessment* (5th ed.). New York: Oxford University Press.

Littman, R., & Takács, Á. (2017). Do all inhibitions act alike? A study of go/no-go and stop-signal paradigms. *PloS one, 12(*10), e0186774. https://doi.org/10.1371/journal.pone.0186774

Livesey, D. J., & Morgan, G. A. (1991). The development of response inhibition in 4- and 5-year-old children. *Australian Journal of Psychology, 43*(3), 133-137. https://doi.org/10.1080/00049539108260137

Luciana, M. (2003). Practitioner review: Computerized assessment of neuropsychological function in children: Clinical and research applications of the Cambridge Neuropsychological Testing Automated Battery (CANTAB). *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *44*(5), 649–663. https://doi.org/10.1111/1469-7610.00152

Luciana, M., & Nelson, C. A. (1998). The functional emergence of prefrontally-guided working memory systems in four-to eight-year-old children. *Neuropsychologia, 36*(3), 273-293. https://doi.org/10.1016/S0028-3932(97)00109-7

Luciana, M., & Nelson, C. A. (2002). Assessment of neuropsychological function through use of the Cambridge Neuropsychological Testing Automated

Battery: Performance in 4- to 12- year-old children. *Developmental Neuropsychology*, *22*(3), 595–624. https://doi.org/10.1207/S15326942DN2203_3

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. *JMIR serious games*, *4*(2), e11. https://doi.org/10.2196/games.5888

Luria, A. R. (1973). The frontal lobes and the regulation of behavior. In K. H. Pribram & A. R. Luria (Eds.), *Psychophysiology of the frontal lobes* (pp. 3-26). Academic Press. https://doi.org/10.1016/B978-0-12-564340-5.50006-8

MacLeod, C. M. (2007). The concept of inhibition in cognition. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in cognition* (p. 3-23). American Psychological Association. https://doi.org/10.1037/11587-001

MacNeill Horton Jr, A. & Soper, H. (2008). The neuropsychology of children's memory. In J. Reed & J. Warner-Rogers (Eds.), *Child neuropsychology: Concepts, theory, and practice* (pp. 218-234). Wiley-Blackwell.

Manly, T., Anderson, V., Crawford, J., George, M., Underbjerg, M., & Robertson, I. H. (2016). The Test of Everyday Attention for Children, second edition (TEA-Ch2). Pearson Clinical Assessment.

McCoy, D. C. (2019). Measuring young children's executive function and self-regulation in classrooms and other real-world settings. *Clinical Child and Family Psychology Review*, *22*(1), 63–74. https://doi.org/10.1007/s10567-019-00285-1

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions. Four general conclusions. *Current directions in psychological science, 21*(1), 8-14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "'Frontal Lobe'" tasks: A latent variable analysis. *Cognitive Psychology*, *41*, 49–100. https://doi.org/10.1006/cogp.1999.0734

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ, 339,* b2535. https://doi.org/10.1136/bmj.b2535

Morooka, T., Ogino, T., Takeuchi, A., Hanafusa, K., Oka, M., & Ohtsuka, Y. (2012). Relationships between the color-word matching Stroop task and the Go/NoGo task: Toward multifaceted assessment of attention and inhibition abilities of children. *Acta Medica Okayama*, *66*(5), 377-386. https://doi.org/10.18926/AMO/49385

Murphy, P. (2017). *Investigating the use of board games as neuropsychological tests with children with acquired brain injury*. (Doctoral dissertation, University of East London). Retrieved from https://repository.uel.ac.uk/item/84vy6

Naglieri, J., A. (2010). *Wechsler nonverbal scale of ability* [PowerPoint presentation]. Retrieved from http://nebula.wsimg.com/1e5d96c8d5e6daac91605ba67fea92da?AccessKeyId=5BBC44D87B24C07C253A&disposition=0&alloworigin=1

Norman, D. G., & Shallice, T. (1986). Attention to action: Willed .and automatic control of behaviour. In R. J. Davidson, G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self regulation* (pp. 1-18). New York, NY: Plenum Press.

Paulus, F. W., Ohmann, S., von Gontard, A., & Popow, C. (2018). Internet gaming disorder in children and adolescents: A systematic review. *Developmental Medicine & Child Neurology*, *60*(7), 645-659. https://doi.org/10.1111/dmcn.13754

Pavitt, A. (2017). *Using a game-like procedure as a new test of problem solving and concept formation in children*. (Doctoral dissertation, University of East London). Retrieved from http://uel.da.ulcc.ac.uk/6736/1/Anne Pavitt.pdf

Petrides, M., & Milner, B. (1982). Deficits on subject-ordered tasks after frontal-and temporal-lobe lesions in man. *Neuropsychologia, 20*(3), 249-262. https://doi.org/10.1016/0028-3932(82)90100-2

Piaget, J. (1936). *Origins of intelligence in the child.* London: Routledge & Kegan Paul.

Pureza, J. R., Jacobsen, G. M., Oliveira, R. G., & Fonseca, R. P. (2011). Relationships between executive functions tasks in late childhood. *Psychology & Neuroscience*, *4*(3), 369–376. https://doi.org/10.3922/j.psns.2011.3.010

Ratiu, P., & Talos, I. F. (2004). Images in clinical medicine: The tale of Phineas Gage. *New England Journal of Medicine, 351*(23), e21. https://doi.org/10.1056/NEJMicm031024

Reitan, R. M. (1992)*. Trail Making Test: Manual for administration and scoring*. New York: Reitan Neuropsychology Laboratory

Reynolds, C. R. (2002). *Comprehensive Trail Making Test.* Austin, TX: Pro-Ed. https://doi.org/10.1177/0734282905282415

Reynolds, C. R., & MacNeill Horton Jr, A. (2008). Assessing executive functions: A life-span perspective. *Psychology in the Schools*, *45*(9), 875–892. https://doi.org/10.1002/pits.20332

Riccio, C. A., Kahn, D. A., Yoon, M., Reynolds, C. R., & Bonura, E. P. (2011). Confirmation of a two-factor model for interpretation of the Comprehensive Trail Making Test with children. *Archives of Clinical Neuropsychology*, *26*(3), 235–239. https://doi.org/10.1093/arclin/acq105

Rizzo, A., Hartholt, A., Grimani, M., Leeds, A., & Liewer, M. (2014). Virtual reality exposure therapy for combat-related posttraumatic stress disorder. *Computer*, *47*(7), 31-37.

Salimpoor, V. N., & Desrocher, M. (2006). Increasing the utility of EF assessment of executive function in children. *Developmental Disabilities Bulletin, 34*(1-2), 15-42.

Senturk, N., Yeniceri, N., Alp, I. E., & Altan-Atalay, A. (2014). An exploratory study on the Junior Brixton Spatial Rule Attainment Test in 6- to 8-year-olds. *Journal of Psychoeducational Assessment*, *32*(2), 123–132. https://doi.org/10.1177/0734282913490917

Sheridan, B. D. (2015). *The modern dragon: Contemporary representations from Tolkein to present* (Doctoral dissertation, University of Waikato). Retrieved from https://researchcommons.waikato.ac.nz/bitstream/handle/10289/9595/thesis.pdf?sequence=3&isAllowed=y

Silver, C. H. (2014). Sources of data about children's executive functioning: Review and commentary. *Child Neuropsychology*, *20*(1), 1–13. https://doi.org/10.1080/09297049.2012.727793

Speed, E., Davison, C., & Gunnell, C. (2016). The anonymity paradox in patient engagement: Reputation, risk and web-based public feedback. *Medical H*umanities, 42(2), 135-140. http://dx.doi.org/10.1136/medhum-2015-010823

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643. https://doi.org/10.1037/h0054651

Stuss, D. T., & Benson, D. F. (1986). *The frontal lobes*. New York: Raven Press.

Taylor, H. G., Albo, V. C., Phebus, C. K., Sachs, B. R., & Bierl, P. G. (1987). Postirradiation treatment outcomes for children with acute lymphocytic leukemia: Clarification of risks. *Journal of Pediatric Psychology*, *12*(3), 395-411. https://doi.org/10.1093/jpepsy/12.3.395

Thorell, L.B., & Nyberg, L. (2008). The Childhood Executive Functioning Inventory (CHEXI): A new rating instrument for parents and teachers. *Developmental Neuropsychology, 33*(4), 536-552. https://doi.org/10.1080/87565640802101516

Timimi, S., & Timimi, L. (2015). The social construction of attention deficit hyperactivity disorder. In M. O’Reilly & J. N. Lester (Eds.), *The Palgrave handbook of child mental health* (pp. 139-157). London: Palgrave Macmillan. https://doi.org/10.1057/9781137428318_8

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, *54*(2), 131–143. https://doi.org/10.1111/jcpp.12001

Unity Technologies. (2018). Unity [Computer software]. Retrieved from https://unity.com/

Unity Technologies. (2020). *Unity asset store*. Retrieved from https://assetstore.unity.com/

Wallisch, A., Little, L. M., Dean, E., & Dunn, W. (2018). Executive function measures for children: A scoping review of ecological validity. *OTJR: Occupation, Participation and Health*, *38*(1), 6–14. https://doi.org/10.1177/1539449217727118

Wasserman, T., & Wasserman, L. D. (2013). Toward an integrated model of executive functioning in children. *Applied Neuropsychology: Child*, *2*(2), 88–96. https://doi.org/10.1080/21622965.2013.748394

Wechsler, D., & Naglieri, J. A. (2006). *Wechsler Non-verbal Scale of Ability*. San Antonio TX: Harcourt Assessment.

Wechsler, D., Kaplan, E., Fein, D., Morris, R., Kramer, J. H., Maerlender, A., & Delis, D. C. (2004). *The Wechsler Intelligence Scale for Children-Fourth Edition Integrated*. San Antonio, TX: The Psychological Corporation.

Weintraub, S., Bauer, P. J., Zelazo, P. D., Wallner-Allen, K., Dikmen, S. S., Heaton, R. K., … Gershon, R. C. (2013). I. NIH Toolbox Cognition Battery (CB): Introduction and pediatric data. *Monographs of the Society for Research in Child Development*, *78*(4), 1–15. https://doi.org/10.1111/mono.12031

Welsh, M., & Peterson, E. (2014). Issues in the conceptualization and assessment of hot executive functions in childhood. *Journal of the International Neuropsychological Society*, *20*(2), 152–156. https://doi.org/10.1017/S1355617713001379

Wiebe, S. A., & McFall, G. P. (2014). Revisiting executive function measurement: Implications for lifespan development. *Measurement: Interdisciplinary Research and Perspectives*, *12*(3), 115–117. https://doi.org/10.1080/15366367.2014.943596

Wiesing, M., Fink, G. R., & Weidner, R. (2020). Accuracy and precision of stimulus timing and reaction times with Unreal Engine and SteamVR. *PloS One*, *15*(4), e0231152. https://doi.org/10.1371/journal.pone.0231152

Williams, D., Yee, N., & Caplan, S. E. (2008). Who plays, how much, and why? Debunking the stereotypical gamer profile. *Journal of computer-mediated communication*, *13*(4), 993-1018. https://doi.org/10.1111/j.1083-6101.2008.00428.x

Willig, C. (2012). Perspectives on the epistemological bases for qualitative research. In H. Cooper (Ed.). *The handbook of research methods in psychology* (pp. 5-21). Washington: American Psychological Association

Willoughby, M. T., & Blair, C. B. (2016). Measuring executive function in early childhood: A case for formative measurement. *Psychological Assessment*, *28*(3), 319–330. https://doi.org/10.1037/pas0000152

Willoughby, M., Holochwost, S. J., Blanton, Z. E., & Blair, C. B. (2014). Executive functions: Formative versus reflective measurement. *Measurement: Interdisciplinary Research and Perspectives*, *12*(3), 69–95. https://doi.org/10.1080/15366367.2014.929453

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, *6*(4), 354–360. https://doi.org/10.1111/j.1750-8606.2012.00246.x

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). II. NIH Toolbox Cognition Battery (CB): Measuring executive function and attention. *Monographs of the Society for Research in Child Development*, *78*(4), 16–33. https://doi.org/10.1111/mono.12032

Zelazo, P. D., Reznick, J. S., & Piñon, D. E. (1995). Response control and the execution of verbal rules. *Developmental Psychology, 31*(3), 508-517. https://doi.org/10.1037/0012-1649.31.3.508

Zhao, Q., Guo, Q., Li, F., Zhou, Y., Wang, B., & Hong, Z. (2013). The Shape Trail Test: Application of a new variant of the trail making test. *PLoS One, 8*(2), e57333. https://doi.org/10.1371/journal.pone.0057333

APPENDICES

APPENDIX A: Cartoon Dragons Used for the Main Characters

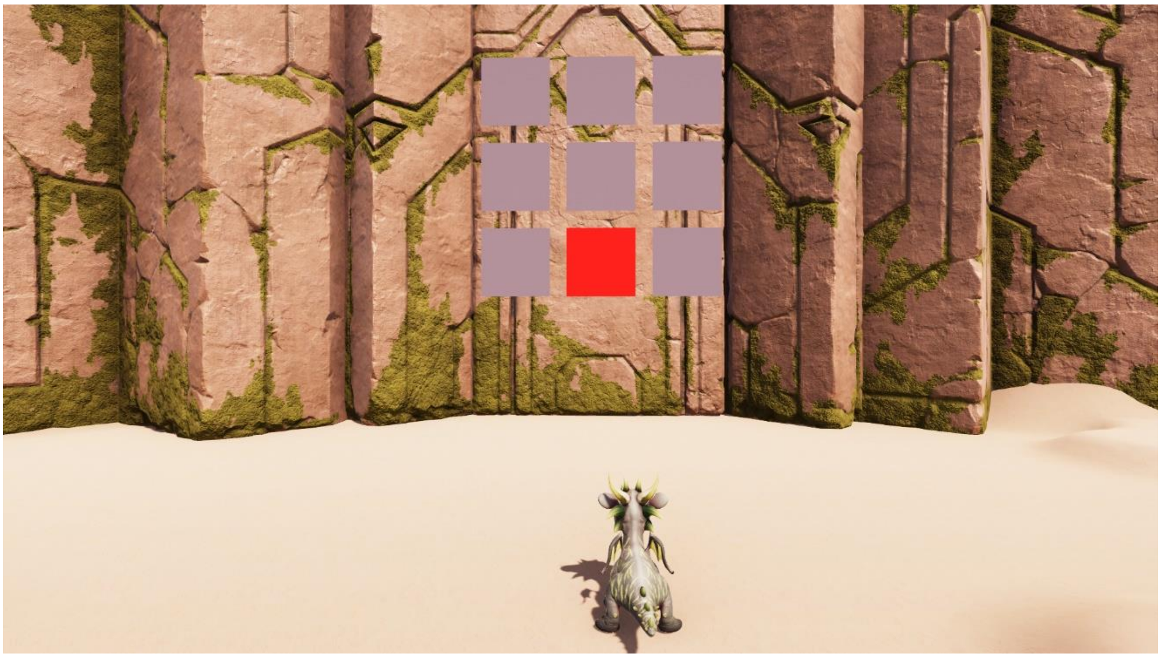
APPENDIX B: **Screenshots of the Three-Dimensional Game Space**

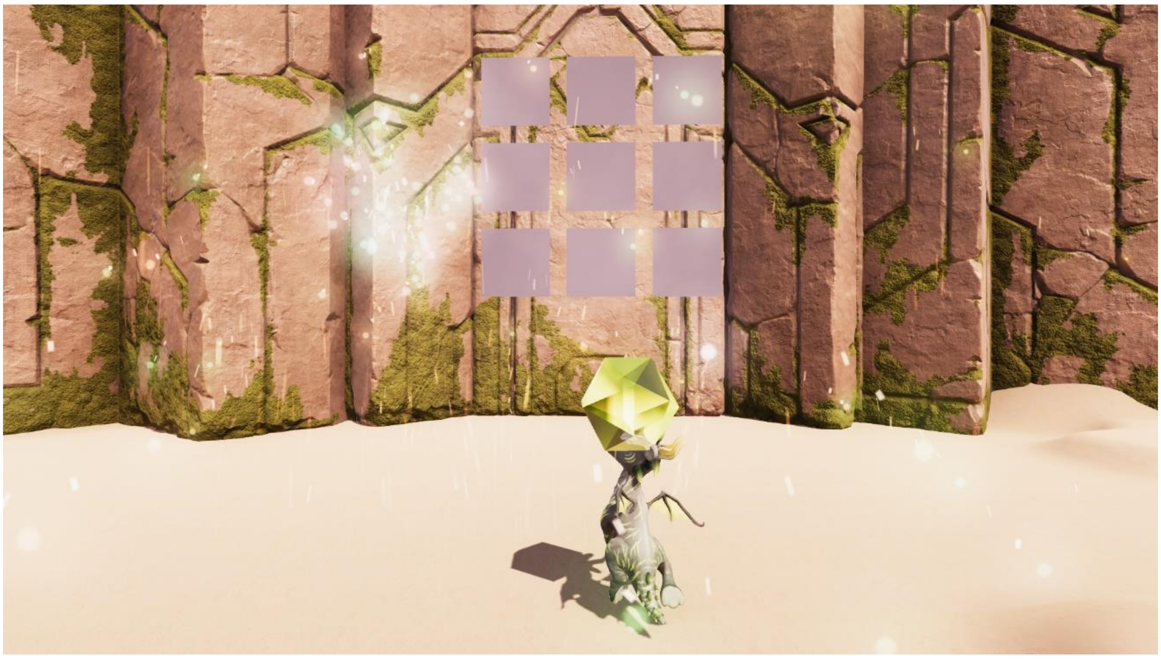
APPENDIX C: **Screenshot of Reward ‘Gems’**

APPENDIX D: **Screenshots of Dragon 'Home'**

APPENDIX E: **Screenshots of Dragon Dash Measure**

APPENDIX F: **Screenshots of Dragon Sequence Measure**





APPENDIX G: **Screenshots of Dragon Hunt Measure**

APPENDIX H: Visual Analogue Scale

Please tell us how you found the task……………..

I enjoyed the task

1 2 3 4 5

Strongly Disagree Disagree Neither Agree nor Disagree Agree Strongly Agree

APPENDIX I: Unity Asset Store Artwork and Acknowledgments to the Authors

Author	Asset Title
256px	Desert Kits 64 Sample
Aquarius Max	Ultimate Fantasy Creator LITE
Baldinoboy	Tropical Forest Pack
Denis Pahunov	Noise Brush
Dustyroom	Free Casual Game SFX Pack
Forst	Conifers [BOTD]
FullTiltBoogie	Starter Particle Pack
Game Sound Solution	Score and Times
GameWarming	Autumn Mountain
InspectorJ Sound Effects	44.1 General Library (Free Sample Pack)
JayAnAm	Low Poly Game Kit
Junnichi Suko	Iguanna
Knyaz	PBR Desert Landscape
Laxer	Modular Wooden Bridge Tiles
Lemuria	Free Desert Plants
MalberS Animations	Little Dragons: Mouse
Manufactura K4	Rock and Boulders 2
Marcos Schultz	MS Advanced Camera Controller
Meshzone3d	Winter Zone Mini
NatureManufacture	Mountain Trees – Dynamic Nature
Playmint	Sets - Gems
ProAssets	Free HDR Sky
Sandro T	Sun Temple
Scrycoast	Cope! Free Skyebox Pack
Shapes	Nature Starter Kit 2
SineVFX	Translucent Crystals
SoundBits	Free Sound Effects
The Tales Factory	Photoscanned MountainsRocks PBR
TurnTheGameOn	Arrow WayPointer
Unity Technologies	3D Game Kit – Environment Pack
Unity Technologies	3D Game Kit – Props Pack
Unity Technologies	Standard Assets
Zosma	Fantasy Free GUI

APPENDIX J: Screenshots of On-screen Instruction Boxes


40
Oivu
To trade with Sizur you need 20
eggs and 20 crystals...
CONTINUE >>

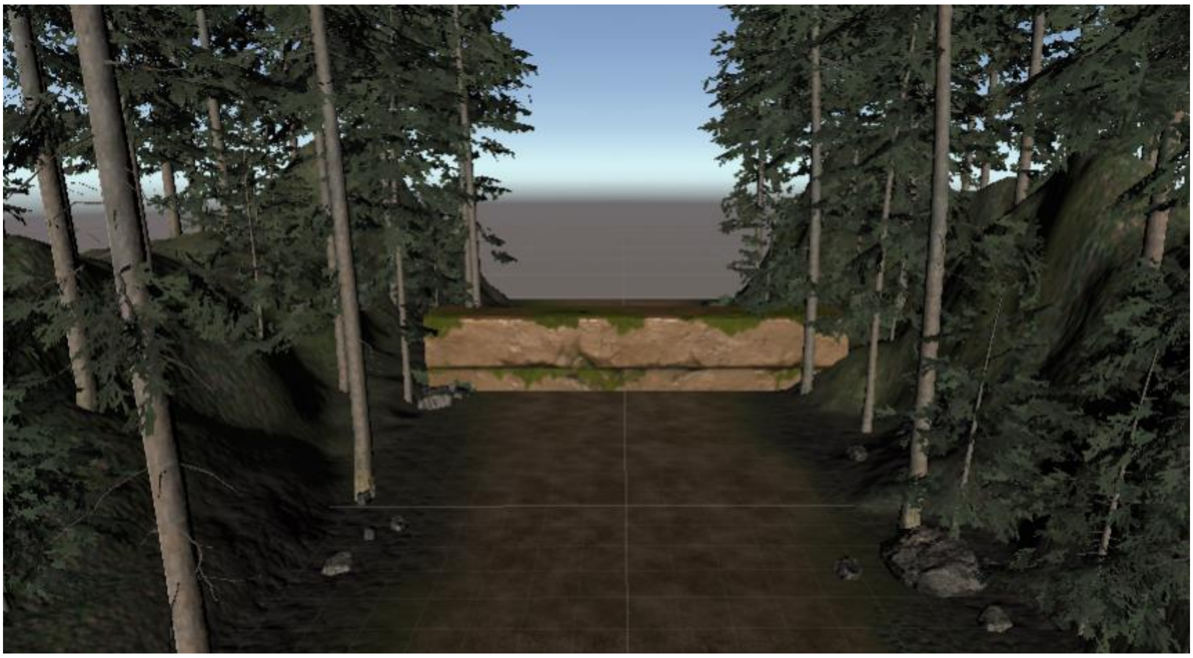
APPENDIX K: Dragon Dash Instructions

- The dragon Sizur has hidden a gem at the end of this path…
- The path has lots of obstacles in the way…
- Avoid the obstacles to get to the end and collect the gem!
- BUT
- Sizur has switched the controls…
- Press the left arrow to move right…
- Press the right arrow to move left…
- Press the up arrow to duck…
- Press the down arrow to jump…
- BUT… watch out!
- When you hear a *beep* the controls will switch back!
- So… left is left…
- Right is right…
- Up is up…
- And down is down…
- Every time you hear the *beep* the controls will switch…
- Good luck!

APPENDIX L: Dragon Dash Left-Hand Block

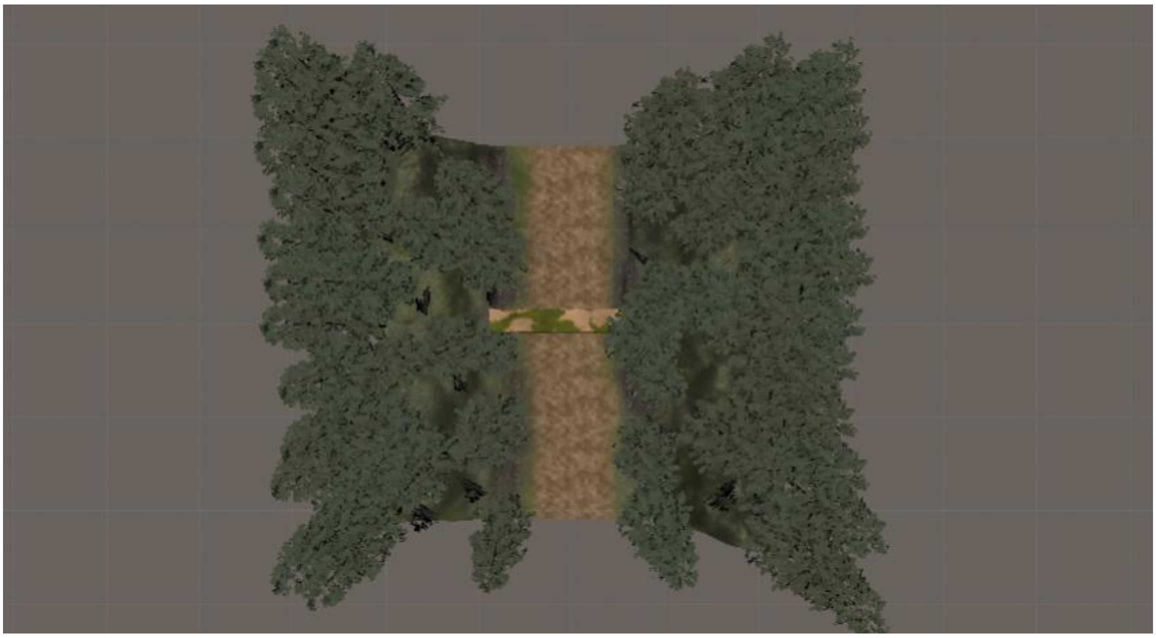
APPENDIX M: Dragon Dash Right-Hand Block

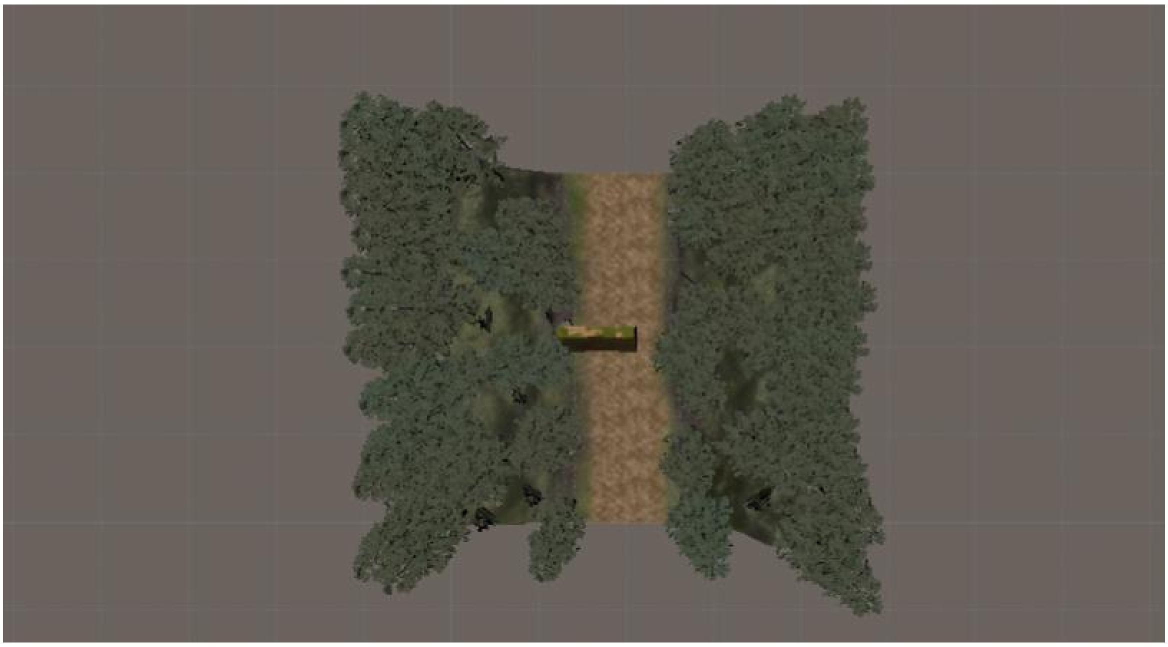
APPENDIX N: Dragon Dash Jumping Block



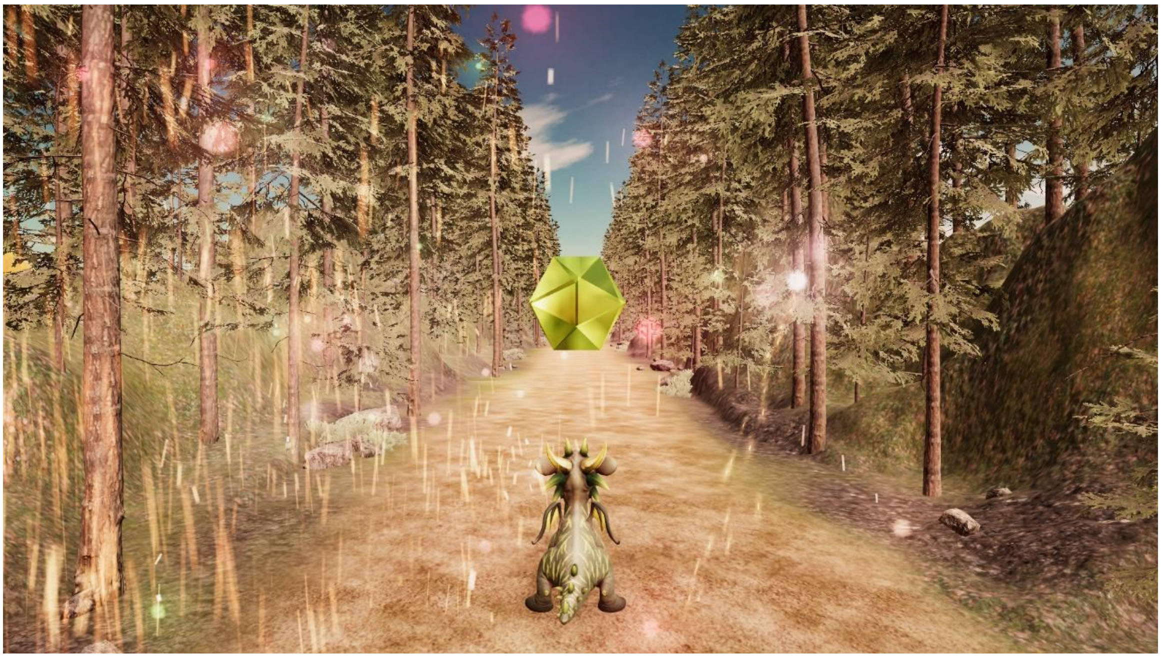
APPENDIX O: Dragon Dash Ducking Block

APPENDIX P: Dragon Dash Platforms





APPENDIX Q: Dragon Dash Celebration Scene



APPENDIX R: Dragon Sequence Instructions

- This gate has nine squares on it…
- To collect the gem you need to copy the squares that light up in the BACKWARDS order…
- First some of the squares will light up…
- Then use the mouse to click the squares that light up in the BACKWARDS order…
- So, if square 1 and 2 light up…
- You click square 2 and then square 1…
- Click the Start button to start…
- Good luck!

APPENDIX S: Dragon Hunt Measure ‘Eggs’

APPENDIX T: Dragon Hunt Measure ‘Crystals’

APPENDIX U: Dragon Hunt Measure Instructions

- Sizur the dragon has the last gem!
- To trade with Sizur you need 20 eggs and 20 crystals…
- There are 10 blue eggs, 10 red eggs, 10 blue crystals and 10 red crystals scattered over the land…
- BUT
- You can only collect them in order…
- Start with the BLUE EGG…
- Then switch COLOUR, to red egg…
- Then switch SHAPE, to red crystal…
- Then switch COLOUR, to blue crystal…
- Keep switching COLOUR and SHAPE, until you have collected them all...Follow the path to find them all…
- Go as quickly as you can without making a mistake…
- Remember to start with the BLUE EGG!

APPENDIX V: UEL Ethical Approval

School of Psychology Research Ethics Committee

NOTICE OF ETHICS REVIEW DECISION

For research involving human participants
BSc/MSc/MA/Professional Doctorates in Clinical, Counselling and Educational Psychology

REVIEWER: [REVIEWER NAME]

SUPERVISOR: **Matthew Jones Chesters**

STUDENT: Jennifer Moynihan

Course: Professional Doctorate in Clinical Psychology

Title of proposed study: The issue of engagement by piloting a novel computerised measure using a game-like protocol.

DECISION OPTIONS:

1. **APPROVED:** Ethics approval for the above named research study has been granted from the date of approval (see end of this notice) to the date it is submitted for assessment/examination.

2. **APPROVED, BUT MINOR AMENDMENTS ARE REQUIRED BEFORE THE RESEARCH COMMENCES** (see Minor Amendments box below)**:** In this circumstance, re-submission of an ethics application is not required but the student must confirm with their supervisor that all minor amendments have been made before the research commences. Students are to do this by filling in the confirmation box below when all amendments have been attended to and emailing a copy of this decision notice to her/his supervisor for their records. The supervisor will then forward the student's confirmation to the School for its records.

3. **NOT APPROVED, MAJOR AMENDMENTS AND RE-SUBMISSION REQUIRED** (see Major Amendments box below)**:** In this circumstance, a revised ethics application must be submitted and approved before any research takes place. The revised application will be reviewed by the same reviewer. If in doubt, students should ask their supervisor for support in revising their ethics application.

DECISION ON THE ABOVE-NAMED PROPOSED RESEARCH STUDY
(Please indicate the decision according to one of the 3 options above)

APPROVED, BUT MINOR AMENDMENTS ARE REQUIRED BEFORE THE RESEARCH COMMENCES

Minor amendments required *(for reviewer)*:

p.3 paragraph two, “measures” should be “measured”.
I wondered why (near the bottom of p.3), ethnic group will be recorded? I think this should be justified in relation to the study’s design?
Bottom of p.3 – is it head teachers, plural, from whom you have obtained permission? Or is there one head teacher for both schools? Clarification would be helpful – I see there is an email from a head teacher asking whether you would like to use Lambeth Academy also – what was your reply, and did the same head teacher, or a different person, then give permission for including that school also?

Major amendments required *(for reviewer):*

Confirmation of making the above minor amendments *(for students)*:

I have noted and made all the required minor amendments, as stated above, before starting my research and collecting data.

Student’s name *(Typed name to act as signature):* Jennifer Moynihan
Student number: U1725761

Date: 01/05/2019

(Please submit a copy of this decision letter to your supervisor with this box completed, if minor amendments to your ethics application are required)

ASSESSMENT OF RISK TO RESEACHER *(for reviewer)*

Has an adequate risk assessment been offered in the application form?

YES / NO

Please request resubmission with an adequate risk assessment

If the proposed research could expose the <u>researcher</u> to any of kind of emotional, physical or health and safety hazard? Please rate the degree of risk:

☐ HIGH

Please do not approve a high risk application and refer to the Chair of Ethics. Travel to countries/provinces/areas deemed to be high risk should not be permitted and an application not approved on this basis. If unsure please refer to the Chair of Ethics.

☐ MEDIUM (Please approve but with appropriate recommendations)

☒ LOW

Reviewer comments in relation to researcher risk (if any).

Reviewer *(Typed name to act as signature)*: [NAME OF REVIEWER]

Date: 29 April 2019

This reviewer has assessed the ethics application for the named research study on behalf of the School of Psychology Research Ethics Committee

RESEARCHER PLEASE NOTE:

For the researcher and participants involved in the above named study to be covered by UEL's Insurance, prior ethics approval from the School of Psychology (acting on behalf of the UEL Research Ethics Committee), and confirmation from students where minor amendments were required, must be obtained before any research takes place.

For a copy of UELs Personal Accident & Travel Insurance Policy, please see the Ethics Folder in the Psychology Noticeboard

UNIVERSITY OF EAST LONDON

School of Psychology

APPLICATION FOR RESEARCH ETHICS APPROVAL

FOR RESEARCH INVOLVING HUMAN PARTICIPANTS

SECTION 1. Your details

1. **Your name:**
   Jennifer Moynihan

2. **Your supervisor's name:**
   Dr Matthew Jones Chesters

3. **Title of your programme:**
   Professional Doctorate in Clinical Psychology

4. **Submission date for your BSc/MSc/MA research:**
   May 2020

5. **Please tick if your application includes a copy of a DBS certificate (see page 3)** ☐

6. **Please tick if your research requires DBS clearance but you are a Prof Doc student and have applied for DBS clearance – or had existing clearance verified – when you registered on your programme (see page 3)** ☑

7. **Please tick if you need to submit a DBS certificate with this application but have emailed a copy to Dr Tim Lomas for confidentiality reasons (Chair of the School Research Ethics Committee) t.lomas@uel.ac.uk** ☐

8. **Please tick to confirm that you have read and understood the British Psychological Society's Code of Ethics and Conduct (2009) and the UEL Code of Practice for Research Ethics (See links on page 1)** ☑

SECTION 2. About your research

9. What your proposed research is about:

Very few measures of executive functioning (EF) have been designed specifically for children. Current measures face several challenges: many are culturally biased (Fernández & Abe, 2018), lack normative data and ecological validity as well as not facilitating the engagement needed to accurately measure EF in children (Anderson 2002b). Game-like procedures have been found to increase engagement (Johann & Karbach, 2018; Józsa, Barrett, & Morgan, 2017). This study aims to address the issue of engagement by piloting a novel computerised measure using a game-like protocol.

The research questions are:

1) Can we pilot a novel computerised game-like measure to test executive functioning by collecting data for typically developing children?
2) Is the novel measure more engaging for children than standardised measures of executive functioning?
3) Does performance on the novel measure correlate with standardised measures of executive functioning?
4) Does performance on the novel measure correlate with a teacher-rated measure of executive functioning?

10. Design of the research:

The study will use a cross-sectional correlational design to investigate relationships between a novel measure, standardised tests and teacher-rated measures.

<u>Novel Measure</u>

The novel measure will include five tests adapted from standardised measures. To increase engagement a young dragon character will be used and the tests will be embedded in a story line.

Flexibility/switching

The dragon must hunt for eggs and gems in two different colours, collecting in alternating order. Performance is measured by completion time and error rate. A ‘practice’ trail will collect processing speed without a switching element.

Abstraction

The dragon must travel from a starting platform to an end platform by jumping on floating platforms. Only one correct platform will be present in each move, requiring the participant to discover the pattern. The pattern will change at intervals, requiring the participant to change their strategy. Performance is measured by error rate.

Working Memory

The dragon is shown a sequence of lights that indicate the code to open a giant door. They must reproduce this pattern (forwards/backwards depending on trial) on a control panel. Performance is measured by total number of components reproduced, and error rate.

Inhibition
The dragon travels along a moving ground surface during which they must avoid numerous obstacles. The obstacles are labelled with arrows and the player must press the opposite key to the direction indicated. Performance is measured via reaction time and error rate.

Inhibitory Control
The dragon must fight 'enemies' by blasting them with fire. The 'enemies' are presented rapidly on screen alongside some 'friends' which must not be hit. Performance is measured by reaction time and error rate.

Standardised Measures
The novel measure will be compared with the comparable standardised measures the tests above were adapted from. Measures are available through the supervisor.

Table 1: *Standardised Tests Included in Study*

Test	Domain
Trail Making Test Part B (Reitan, 1992)	Flexibility/switching
The Brixton Spatial Anticipation Test (Burgess & Shallice, 1997)	Abstraction
The WMC-IV Spatial Addition (Wechsler, 2010)	Working Memory
The Colour Word Interference Test (Delis, Kaplan & Kramer, 2001)	Inhibition
Sustained Attention to Response Task (Robertson, Manly, Andrade, Baddeley & Yiend, 1997)	Inhibitory Control

Functional Measures
The *Childhood Executive Functioning Inventory* (CHEXI, Thorell & Nyberg, 2008) will be used to measure EF in every-day behaviour. It is suitable for rating by teachers and freely available.

A visual analogue tool will be used for participants to rate how much they enjoyed each task

11. Recruitment and participants (Your sample):

Participants will be typically developing children between 11-12 years old. A limited age range was chosen as creating an age-appropriate game across childhood is beyond the scope of the project and not yet clinically indicated. 11 and 12-year olds were chosen as this is the age range in which it is believed EF becomes relatively established (Goldstein & Naglieri, 2014).

As the study aims to investigate typically developing children, individuals with significant learning impairments will be excluded. Participants will also need to be able to speak and read English. Both males and females will be recruited, aiming for equal numbers. Ethnic group and whether English is an additional or primary language will be recorded so as to provide some information to indicate whether the test is "culture fair".

The study aims to recruit 45-50 participants. Participants will be recruited from two secondary schools in London, permission has been obtained from the head teacher and he is the head teacher of both schools.

12. Measures, materials or equipment:
Please refer to section 10 for measures. In addition, the study will require access to the Unity game development platform, access is free for projects of this size. Access to a quiet room within the schools and a password-protected computer is also required.

13. If you are using copyrighted/pre-validated questionnaires, tests or other stimuli that you have not written or made yourself, are these questionnaires and tests suitable for the age group of your participants?

YES

14. Outline the data collection procedure involved in your research:
Consent will be requested from parents/guardians and from the child themselves. Children will be provided with the opportunity to discuss and ask questions about the study. If consent is given, the study will continue.

Demographic information will be requested: D.O.B., gender, ethnicity and whether English is a primary or additional language.

The participants will then be asked to complete the standardised measures and play the novel game-like measure, the order of which will be counter-balanced. The administration will take place in a quiet room in the school.

After each task, the participants will rate how enjoyable they found the task.

It is estimated that the tasks will take no longer than 1 hour, participants will be offered a break half-way through.

Following completion of the study participants will be reminded of what the study was about. If they would like more information about the study, they will be able to speak to the researcher or their teacher.

Finally, teachers will be approached to complete the CHEXI.

SECTION 3. Ethical considerations

15. Fully informing participants about the research (and parents/guardians if necessary):
Parents/guardians will receive an information letter about the study and inviting their children to participate. Participants will receive an information sheet that is developmentally appropriate.

16. Obtaining fully informed consent from participants (and from parents/guardians if necessary):

Consent will be requested from parents/guardians via an information form and consent form.

Following this consent would also be obtained from the participants. A developmentally appropriate information and consent form would be provided and the researcher would be available to answer any questions.

17. Engaging in deception, if relevant:

This study does not involve deception.

18. Right of withdrawal:

Parents/guardians and participants will be informed of their right to withdraw from the study at any time without having to provide a reason and without any disadvantage to themselves. They will also be informed that they have the right to withdraw their data from the study and that it will not be used in the final analysis if a request to withdraw the information has been made within three weeks of their participation.

This information will be written in the information sheets for parents/guardians and the participants.

19. Will the data be gathered anonymously?

NO

20. If NO what steps will be taken to ensure confidentiality and protect the identity of participants?

Participant's data will be anonymised by allocating a participant code to corresponding data. The participant code will be used instead of names in the database. Participant names and codes will be stored in a separate password-protected file.

All data, including identifying information will be securely stored in password-protected files in accordance with GDPR regulations.

At the end of the study participant names and associated codes will be destroyed. The remaining data will be help for up to two years to support publication of the results.

21. Will participants be paid or reimbursed? NO

<u>SECTION 4. Other permissions and ethical clearances</u>

22. Research involving the NHS in England

Is HRA approval for research involving the NHS required? NO

Will the research involve NHS employees who will not be directly recruited through the NHS and where data from NHS employees will not be collected on NHS premises?

NO

If you work for an NHS Trust and plan to recruit colleagues from the Trust will permission from an appropriate member of staff at the Trust be sought and is a copy of this permission (can be an email from the Trust) attached to this application?

N/A

23. Permission(s) from an external institution/organisation (e.g. a school, charity, workplace, local authority, care home etc.)?
Permission to recruit from the schools has been obtained from the head teacher and attached to this document.

Is permission from an external institution/organisation/workplace required? YES

If YES please give the name and address of the institution/organisation/workplace:

[SCHOOL NAME]

[SCHOOL ADDRESS]

24. Is ethical clearance required from any other ethics committee?
NO
If YES please give the name and address of the organisation:

Has such ethical clearance been obtained yet? N/A

If NO why not?
If YES, please attach a scanned copy of the ethical approval letter. A copy of an email from the organisation confirming its ethical clearance is acceptable.

<u>SECTION 5. Risk Assessment</u>
Any concerns about the safety of the participants or researchers will be reported to the supervisor as soon as possible.

25. Protection of participants:
It is not anticipated that the study would create significant physical, emotional or psychological harm. However, all participants will be monitored through-out the study for signs of distress and/or fatigue. A break will be offered half-way through the study. After each test the participants will indicate how much they enjoyed / did not enjoy the test. After completion of the study, verbal feedback will be requested as to how they experienced it. The participants will be offered the opportunity to talk to the researcher or their teacher about the study.

If the tests indicated any significant harm to the participants this would be discussed with the supervisor as soon as possible and consent requested to share this with the parents/guardians and school.

26. Protection of the researcher:

There are no known risks to the researcher. If any risks were to occur, the supervisor would be informed as soon as possible. Local protocol would also be followed within the school if risks were identified.

27. Debriefing participants:

Following completion of the study, participants will be reminded of what the study was about. They will also be offered the opportunity to talk to the researcher or their teacher about any questions they may have about the study.

28. Other:

N/A

29. Will your research involve working with children or vulnerable adults?*
YES

If YES have you obtained and attached a DBS certificate?

DBS has been obtained through the Professional Doctorate in Clinical Psychology.

If your research involves young people under 16 years of age and young people of limited competence will parental/guardian consent be obtained.

YES

If NO please give reasons. (Note that parental consent is always required for participants who are 16 years of age and younger)

30. Will you be collecting data overseas? NO

If YES in what country or countries (and province if appropriate) will you be collecting data?

N/A

SECTION 6. Declarations

Declaration by student:

I confirm that I have discussed the ethics and feasibility of this research proposal with my supervisor.

Student's name: Jennifer Moynihan
Student's number: 1725761
Date: 28/02/2019

Supervisor's declaration of support is given upon their electronic submission of the application

University of East London Psychology

REQUEST FOR TITLE CHANGE TO AN ETHICS APPLICATION

FOR BSc, MSc/MA & TAUGHT PROFESSIONAL DOCTORATE STUDENTS

Please complete this form if you are requesting approval for proposed title change to an ethics application that has been approved by the School of Psychology.

By applying for a change of title request you confirm that in doing so the process by which you have collected your data/conducted your research has not changed or deviated from your original ethics approval. If either of these have changed then you are required to complete an Ethics Amendments Form.

HOW TO COMPLETE & SUBMIT THE REQUEST

1. Complete the request form electronically and accurately.
2. Type your name in the 'student's signature' section (page 2).
3. Using your UEL email address, email the completed request form along with associated documents to: Psychology.Ethics@uel.ac.uk
4. Your request form will be returned to you via your UEL email address with reviewer's response box completed. This will normally be within five days. Keep a copy of the approval to submit with your project/dissertation/thesis.

REQUIRED DOCUMENTS

1. A copy of the approval of your initial ethics application.

Name of applicant: Jennifer Davis (née Moynihan)
Programme of study: Professional Doctorate in Clinical Psychology
Name of supervisor: Dr Matthew Jones Chesters

Briefly outline the nature of your proposed title change in the boxes below

Proposed amendment	Rationale
Old Title: The issue of engagement by piloting a novel computerised measure using a game-like protocol. **New Title:** Using a Novel Game-Like Computerised Measure to Test Executive Functioning in Children	The old title did not match the registered title on PHD manager. The new title is the registered title of the project on PHD manager.

Please tick	YES	NO
Is your supervisor aware of your proposed amendment(s) and agree to them?	X	
Does your change of title impact the process of how you collected your data/conducted your research?		X

Student's signature (please type your name): JENNIFER DAVIS (Moynihan)

Date: 17/03/2020

TO BE COMPLETED BY REVIEWER		
Title changes approved	APPROVED	
Comments N/A		

Reviewer: [REVIEWER NAME]

Date: 23rd March 2020

APPENDIX W: Parent/ Guardian Information Form

INFORMATION SHEET FOR PARENTS/GUARDIANS

The children in Year 7 are being invited to participate in a research study. Before deciding whether you would like your child to participate, this sheet provides information about the study and what participation would involve.

Who am I?
My name is Jenny Davis. I am a Trainee Clinical Psychologist studying at the University of East London. As part of my training I am conducting the research your child is being invited to participate in.

What is the research?
The aim of the research is to develop a new measure to test children's ability to plan, adjust and organise their thinking and behaviour. These skills are needed for many tasks in every-day life and school, however, there is a lack of tests designed specifically for children. We have developed a computer-game designed to test these abilities. The aim is to find out whether children find this game more engaging than traditional measures. If children do find it more engaging it could help us measure these skills more accurately. This research project has been approved by the School of Psychology Research Ethics Committee. This means that the research follows the standard of research ethics set by the British Psychological Society.

What will happen if I allow my child to take part?
The study will take place in a quiet room in your child's school. The researcher will speak with your child about the study and they can ask any questions they may have. If they would like to participate they will fill in a consent form.

The study would involve your child completing some pen and paper measures and the computer game. We would then ask your child to tell us how enjoyable these were. It will take approximately 1 hour and your child will be offered a break in the middle.

Finally, we would ask one of your child's teachers to fill in a brief questionnaire about your child's ability to plan, adjust and organise their thinking and behaviours. The aim of this is to find out whether the measures are related to real-life strengths and/or difficulties.

Does my child have to take part?

No, it is your choice whether you would like you child to take part or not. The study is independent of your child's school and participation is not linked to their education. If you decide you do not want your child to participate you do not have to give a reason.

You can withdraw your consent for your child to participate at any time, even if you initially provided consent or they have already completed the study. Your child can also tell us at any time if they would like to withdraw, if this were to happen you would be notified. When consent is withdrawn all data associated with your child would be withdrawn and not used in the final analysis. Please notify your child's teacher within three weeks of participation to prevent their data being used in the final analysis.

What will happen to the information that your child provides?
Your child's privacy and safety will be respected at all times. All the information collected will be anonymised, kept confidential and securely stored. Your child's consent form will be kept separately from the rest of the data. The researcher and the project supervisor are the only people who will see your child's responses. At the end of the study any information with your child's name on it (e.g. consent forms) will be destroyed. Anonymised information may be kept for up to two years.

What will happen to the results of the study?
The results of the study will be written up into a report and submitted as part of a doctorate in Clinical Psychology. At a later stage the results may also be published as a journal article. Your child would not be identifiable in either of these reports.

Contact Details
If you would like further information about my research or have any questions or concerns, please do not hesitate to get in contact:

Jenny Davis (Trainee Clinical Psychologist), U1725761@uel.ac.uk

If you have any questions or concerns about how the study has been conducted, please contact my supervisor Name: Dr Matthew Jones-Chesters, School of Psychology, University of East London, Water Lane, London E15 4LZ. 020 8223 4082 Email: m.h.jones-chesters@uel.ac.uk

or

Chair of the School of Psychology Research Ethics Sub-committee: Dr Tim Lomas, School of Psychology, University of East London, Water Lane, London E15 4LZ (Email: t.lomas@uel.ac.uk).

APPENDIX X: Parent/ Guardian Consent Form

UEL
University of
East London

UNIVERSITY OF EAST LONDON

Consent to participate in a research study

Study Title: Using a novel game-like computerised measure to test executive functioning in children.

I have the read the information sheet relating to the above research study and have been given a copy to keep. The nature and purposes of the research have been explained to me, and I have had the opportunity to discuss the details and ask questions about this information. I understand what is being proposed and the procedures in which my child will be involved in have been explained to me.

I understand that my child's involvement in this study, and particular data from this research, will remain strictly confidential. Only the researcher(s) involved in the study will have access to identifying data. It has been explained to me what will happen once the research study has been completed.

I hereby freely and fully consent for my child to participate in the study which has been fully explained to me. Having given this consent I understand that I have the right to withdraw my child from the study at any time without disadvantage to myself or my child and without being obliged to give any reason. I also understand that should I withdraw my child, any data that they have provided will be withdrawn from the study and not used in the final analysis.

I consent ☐ **OR** **I do not consent** ☐

Name of parent/guardian

……………………………………………………………………………………

Name of child

…………………………………………………………………………………..

Parent/guardian's signature

…………………………………………………………………………………..

Researcher's Name

…………………………………………………………………………………..

Researcher's Signature

……………………………………………………………………………………

Date: ……………………..…….

APPENDIX Y: Child Information and Consent Form

UNIVERSITY OF EAST LONDON

Information and consent to participate in a research study

Study Title: Using a novel game-like computerised measure to test executive functioning in children.

I am studying the strategies people use to solve problems whilst playing a computer game, compared to some spoken and pen and paper tasks. I am asking students in year 7 to take part in the research project.

If you decide to take part, I will ask you to play a short computer game and pen and paper tasks. The computer game will involve completing some puzzles to complete the levels and finish the game.

After the computer game I will ask you to try four more tasks. The first one involves using a pencil to join up some symbols, the second involves reading some simple words out-loud, the third involves spotting patterns and the last involves remembering some numbers.

It is your choice whether you want to take part or not, and nothing will happen if you decide not to take part. If you change your mind after you have started you can stop at any time and you don't have to say why.

If you would like to find out more about the study you can ask me or your teacher.

Would you like to take part in the study? Yes □ No □

Participant's Name
………………………………………………………………………………………….

Researcher's Name
…………………………………………………………………………………………..

Researcher's Signature
……………………………………………………………………………………………

Date: ……………………………..…….